


```

111  :::::::::::::::::::::::::::::::::::::::::::: |||::
99 HisPhePheProLysSerAsnLeuGluHisHisHisHisHisValG1 115
10 C 10
:
115 u 115

```

seq_name: /SIDSI/gcgdata/hold-gene-seq/geneseq-emb1/AA1992.DAT:AA27492

seq_documentation_block:

ID AAR27492 standard; Protein; 147 AA.

XX

AC AAR27492;

XX

DT 05-MAR-1993 (first entry)

XX

DE Fusion protein expressed by pMTNF-MPH.

XX

KW Mycobacterium; Crohn's disease; John's disease; cattle; human;

KW M. paratuberculosis; TNF.

XX

OS Mycobacterium paratuberculosis.

XX

Key Location/Qualifiers

FT Peptide

FT Peptide

FT Peptide

FT Protein

FT Protein

FT Protein

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

60 CATGCTTACCAAAAATAAGCTACACCATCACCATCACCATTAAAGTCGA 11
111  :::::::::::::::::::::::::::::::::::::::::::: |||::
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHisHisHisHisHis 36
10 CCTCTCGCT 1
111  |||
36 pProGlyPro 39

```

seq_name: /SIDSI/gcgdata/hold-gene-seq/geneseq-emb1/AA2000.DAT:AA43894

seq_documentation_block:

ID AAY43894 standard; Protein; 318 AA.

XX

AC AAY43894;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

alignment_scores:
Quality: 64.50 Length: 20
Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x AAR27492 ..
Align seg 1/1 to: AAR27492 from: 1 to: 147

alignment_scores:
Quality: 64.50 Length: 20
Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

Example 1: Fig 2-2; 66pp; English.

The present sequence represents a fusion protein comprising a Hepatitis C virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3 protease protein) in the presence of reducing agent on the solid phase. Use of the reducing agent and purification of the antigen using CC sulphonation and desulphonation steps increases its reactivity to HCV antibodies and enables an earlier detection of HCV infection. The assay is used for detecting antibodies raised against the HCV antigen. The polypeptides are used for preventing and treating HCV infection. The polypeptides are also used for diagnosing hepatitis infection. The antibodies to these polypeptides are used for providing passive vaccination.

New hepatitis C-virus polypeptide used for treating the infection -

Example 1: Fig 2-2; 66pp; English.

The present sequence represents a fusion protein comprising a Hepatitis C virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3 protease protein) in the presence of reducing agent on the solid phase. Use of the reducing agent and purification of the antigen using CC sulphonation and desulphonation steps increases its reactivity to HCV antibodies and enables an earlier detection of HCV infection. The assay is used for detecting antibodies raised against the HCV antigen. The polypeptides are used for preventing and treating HCV infection. The polypeptides are also used for diagnosing hepatitis infection. The antibodies to these polypeptides are used for providing passive vaccination.

Sequence 318 AA;

alignment_scores:
Quality: 64.50 Length: 20
Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
 US-09-674-779-4/rev x AAY43898 ..
 Align seg 1/1 to: AAY43898 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACACATCACCACCATTAAGTCGA 11
 ||| :::: ::||| ||||| ||||| |||||
 21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...Valas 36

10 CCTCTGCCT 1
 |||| |||
 36 pProGlyPro 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT.AAY43899

seq_documentation_block:
 ID_ AAY43899 standard; Protein; 318 AA.
 XX AC AAY43899;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the mTNFh6NS3 Type 2b fusion protein.
 XX
 KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Region 1..40
 FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"
 XX
 PN WO9954735-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 15-APR-1999; 99WO-EP02547.
 XX
 PR 17-APR-1998; 98EP-0870087.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;
 XX
 DR WPI; 2000-013283/01.
 DR N-PSDB; AAZ36168.
 XX
 PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 6; Fig 7-2; 66pp; English.
 XX
 CC The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV

CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
 US-09-674-779-4/rev x AAY43899 ..
 Align seg 1/1 to: AAY43899 from: 1 to: 318

60 CATGCTTACCAAAATAAGCTACACATCACCACCATTAAGTCGA 11
 ||| :::: ::||| ||||| ||||| |||||
 21 HisGlnValGluGlnGlyIleHisHisHisHisHisHis...Valas 36

10 CCTCTGCCT 1
 |||| |||
 36 pProGlyPro 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-embl/AA2000.DAT.AAY43900

seq_documentation_block:
 ID_ AAY43900 standard; Protein; 318 AA.
 XX AC AAY43900;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the mTNFh6NS3 Type 2c fusion protein.
 XX
 KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.
 XX
 OS Synthetic.
 OS Hepatitis C virus.
 XX
 FH Key Location/Qualifiers
 FT Region 1..40
 FT /note= "these residues represent the non-NS3
 FT sequence, which is the mTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"
 XX
 PN WO9954735-A1.
 XX
 PD 28-OCT-1999.
 XX
 PF 15-APR-1999; 99WO-EP02547.
 XX
 PR 17-APR-1998; 98EP-0870087.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;
 XX
 DR WPI; 2000-013283/01.
 DR N-PSDB; AAZ36169.
 XX
 PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 7; Fig 8-2; 66pp; English.
 XX
 CC The present sequence represents a fusion protein comprising a Hepatitis C

CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX
 SQ Sequence 318 AA;

alignment_scores:
 Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-674-779-4/rev x AAY43900 ..

Align seg 1/1 to: AAY43900 from: 1 to: 318

```

60 CATGCTTACCAAAATAAGCTACACCATCACCATCACCATTAAAGTCGA 11
   |||   ::   ::   ::   ::   ::   ::   ::   ::   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHisHisHis...Valas 36
   |||   |||
10 CCTCTGCGCT 1
   |||   |||
36 pProGlyPro 39

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: AAY43896

seq_documentation_block:

ID AAY43896 standard; Protein; 326 AA.

XX AAY43896;

XX 11-FEB-2000 (first entry)

DE Amino acid sequence of the MTNHF6NS3 Type 3a clone 21 fusion protein.

KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.

XX Synthetic.

OS Hepatitis C virus.

XX Key Location/Qualifiers
 XX Region 1..48

FT /note= "these residues represent the non-NS3
 FT sequence, which is the MTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.

XX Maertens G, Louwagie J, Bosman A, Sablon E, Zrein M;

XX WPI: 2000-013283/01.

XX N-PSDB; AA236165.

XX

PT New hepatitis C-virus polypeptide used for treating the infection -
 XX
 PS Example 4; Fig 4-2; 66pp; English.
 XX

CC The present sequence represents a fusion protein comprising a Hepatitis C
 CC virus (HCV) NS3 protein. The NS3 polypeptides are used in a solid phase
 CC immunoassay comprising a HCV antigen (preferably a NS3 helicase or NS3
 CC protease protein) in the presence of reducing agent on the solid phase.
 CC Use of the reducing agent and purification of the antigen using
 CC sulphonation and desulphonation steps increases its reactivity to HCV
 CC antibodies and enables an earlier detection of HCV infection. The assay
 CC is used for detecting antibodies raised against the HCV antigen. The
 CC polypeptides are used for preventing and treating HCV infection. The
 CC polypeptides are also used for diagnosing hepatitis infection. The
 CC antibodies to these polypeptides are used for providing passive
 CC vaccination.
 XX

SQ Sequence 326 AA;

alignment_scores:

Quality: 64.50 Length: 20
 Ratio: 4.300 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-674-779-4/rev x AAY43896 ..

Align seg 1/1 to: AAY43896 from: 1 to: 326

```

60 CATGCTTACCAAAATAAGCTACACCATCACCATCACCATTAAAGTCGA 11
   |||   ::   ::   ::   ::   ::   ::   ::   ::   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHisHisHis...Valas 36
   |||   |||
10 CCTCTGCGCT 1
   |||   |||
36 pProGlyPro 39

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT: AAY43897

seq_documentation_block:

ID AAY43897 standard; Protein; 326 AA.

XX AAY43897;

XX 11-FEB-2000 (first entry)

DE Amino acid sequence of the MTNHF6NS3 Type 3a clone 32 fusion protein.

KW HCV; NS3 helicase; HCV subtype 1a; HCV subtype 1b; HCV infection;
 KW solid phase immunoassay; HCV antigen; NS3 protease; HCV antibody;
 KW passive vaccination.

XX Synthetic.

OS Hepatitis C virus.

XX Key Location/Qualifiers
 XX Region 1..48

FT /note= "these residues represent the non-NS3
 FT sequence, which is the MTNF fusion partner,
 FT the hexahistidine tag and part of the
 FT multilinker"

XX WO9954735-A1.

XX 28-OCT-1999.

XX 15-APR-1999; 99WO-EP02547.

XX 17-APR-1998; 98EP-0870087.

XX (INNO-) INNOGENETICS NV.


```

XX PR 26-JAN-1996; 96EP-0870006.
XX PA (INNO-) INNOGENETICS NV.
XX PI Jacobs D, Saman E, Van Heuverswyn H;
XX XX WPI; 1997-393688/36.
XX DR
XX XX New isolated Toxoplasma gondii antigen Tg20 - used to develop
XX PT products for the diagnosis and prevention of T. gondii infection in
XX PT mammals
XX XX
XX PS Example 3; Fig 4; 68pp; English.
XX CC This sequence comprises a fusion protein composed of the mouse
XX CC tumour necrosis factor leader peptide, a polyhistidine tag, and
XX CC antigen Tg20 (see AAW26692) of Toxoplasma gondii. A vector
XX CC containing DNA encoding the fusion protein was used to transform
XX CC Escherichia coli MN1061 host cells for production of recombinant
XX CC Tg20. Antigen Tg20 polypeptides and immunodominant peptides (see
XX CC AAW26693-97) can be used in claimed methods for the diagnosis and
XX CC prevention of T. gondii infection in mammals.
XX SQ Sequence 272 AA;

alignment_scores:
    Quality: 61.50      Length: 18
    Ratio: 4.393       Gaps: 1
    Percent Similarity: 77.778      Percent Identity: 55.556

alignment_block:
US-09-674-779-4/rev x AAW26698 ..
Align seg 1/1 to: AAW26698 from: 1 to: 272

50 CATGCTTACCAAAATAAGCTACACATCACCATCACCATTAAAGTCGA 11
|||:::|||||:|||||:|||||:|||||:|||||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...ValAs 36
10 CCCT 7
|||
36 pPro 37

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AA67271
seq_documentation_block:
ID AAG67271 standard; protein; 282 AA.
XX AC AAG67271;
XX XX
XX DT 13-NOV-2001 (first entry)
XX DE Amino acid sequence of the S.aureus NAD synthetase (NadE).
XX KW NAD synthetase; NadE; protein co-ordinate data.
XX OS Staphylococcus aureus.
XX XX
XX PN WO20015443-A1.
XX XX
XX PD 02-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US02913.
XX XX
XX PR 31-JAN-2000; 2000US-0179261.
XX XX
XX PA (PHAA ) PHARMACIA & UPJOHN CO.
XX XX
XX PI Benson TE, Prince DB;
XX XX
XX DR WPI; 2001-488798/53.

```

```

XX A molecule comprising an S.aureus NAD synthetase or NAD synthetase-like
XX PT substrate binding pocket with defined structure coordinates for
XX PT identifying inhibitors of NAD synthetase for rational drug design -
XX PS Claim 42; Fig 9; 940pp; English.
XX XX
XX CC The present sequence represents a NAD synthetase, designated NadE,
XX CC which is isolated from Staphylococcus aureus. The NadE polypeptide is
XX CC defined by a set of points with a root mean square deviation of less
XX CC than about 1.1 Angstrom from points representing the backbone of the
XX CC amino acids as represented by structure coordinates fully defined in
XX CC the specification. The protein-coordinate data for NadE is useful for
XX CC identifying and making an inhibitor of NadE activity. It is also useful
XX CC for drug discovery using rational drug design. The data is useful for
XX CC solving the structure of other crystal forms of S.aureus NadE or
XX CC S. aureus NadE complexes.
XX SQ Sequence 282 AA;

alignment_scores:
    Quality: 61.00      Length: 11
    Ratio: 6.100       Gaps: 0
    Percent Similarity: 90.909      Percent Identity: 72.727

alignment_block:
US-09-674-779-4/rev x AAG67271 ..
Align seg 1/1 to: AAG67271 from: 1 to: 282

51 CCAAAATAAGCTACACATCACCATCACCAT 19
|||||:|||||:|||||:|||||:|||||
272 ProLysSerArgSerHisHisHisHisHis 282

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:AAO11238
seq_documentation_block:
ID AAO11238 standard; Protein; 109 AA.
XX AC AAO11238;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 25130.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX XX
XX PN WO200164835-A2.
XX XX
XX PD 07-SEP-2001.
XX XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX XX
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX XX
XX PI Tang YT, Liu C, Drmanac RT;
XX XX
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AA191169.
XX XX
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -

```

PS Claim 20; SEQ ID NO 25130; 1399pp + Sequence Listing; English.
 XX The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA00010-AA013910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 109 AA;

alignment_scores:
 Quality: 60.00 Length: 14
 Ratio: 5.455 Gaps: 0
 Percent Similarity: 78.571 Percent Identity: 57.143

alignment_block:
 US-09-674-779-4/rev x AA011238 ..

Align seg 1/1 to: AA011238 from: 1 to: 109

60 CATGCTTACCAAAATAAGCTACCATCACCATCACCAT 19
 |||::: ::|||::: ||||| ||||| |||||
 59 HisLeu***AsnArgAsnSerHisHisHisHisHis 72

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:AA44485

seq_documentation_block:

ID AA44485 standard; Protein; 206 AA.

XX AC AA44485;

XX 27-MAR-2000 (first entry)

DE Human Interleukin 17C with C-terminal Gly(His)8 tag, IL-17C.his.

XX Interleukin; IL-17C.his: PRO1122 polypeptide; clone DNA62377-1381-1;
 KW immunoprecipitation; IL-17 receptor extracellular domain; IL-17R ECD;
 KW cytokine IL-17; hybridisation probe; antagonist; Gly(His)8 tag; agonist;
 KW degenerative cartilaginous disorder; diagnose; therapy.
 XX

OS Homo sapiens.

XX Key Location/Qualifiers

FH Peptide 1..18

FT /label= Signal_peptide

FT Protein 19..197

FT /label= Mature_IL-17C_polypeptide
 FT /note= "Used to treat degenerative cartilaginous
 disorder"

FT Misc-difference 109

FT /note= "Conserved Trp residue"

FT Misc-difference 129

FT /note= "Conserved Cys residue"

FT Misc-difference 134

FT /note= "Conserved Cys residue"

FT Misc-difference 163

FT /note= "Conserved Cys residue"

FT Misc-difference 189

FT /note= "Conserved Cys residue"

FT Misc-difference 191

FT /note= "Conserved Cys residue"

FT Misc-difference 198..206

FT /note= "C-terminal Gly(His)8 tag"

FT

XX WO9960127-A2.
 PN
 XX
 PD 25-NOV-1999.
 XX
 PF 14-MAY-1999; 99WO-US10733.
 XX
 PR 15-MAY-1998; 98US-0085579.
 XX
 PR 23-DEC-1998; 98US-0113621.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Chen J, Filvaroff E, Goddard A, Gurney AL, Li H, Wood WI;
 XX
 DR WPT; 2000-116314/10.
 XX
 XX New polypeptides designated PRO1031 and PRO1122 used to treat a
 PT degenerative cartilaginous disorder -
 XX
 PS Example 11; Page 138-139; 141pp; English.

XX The present sequence is the human PRO1122 polypeptide, with a C-terminal
 CC Gly(His)8 tag, IL-17C.his, derived from the clone DNA62377-1381-1.
 CC This sequence is used in a competitive binding experiment for the
 CC immunoprecipitation of IL-17 receptor extracellular domain (IL-17R ECD).
 CC The entire coding region of IL-17C can be used as hybridisation probe.
 CC The PRO1122 polypeptide, agonist or antagonist, is used to diagnose and
 CC treat a degenerative cartilaginous disorder.
 XX
 SQ Sequence 206 AA;

alignment_scores:
 Quality: 60.00 Length: 13
 Ratio: 5.455 Gaps: 0
 Percent Similarity: 84.615 Percent Identity: 69.231

alignment_block:

US-09-674-779-4/rev x AA44485 ..

Align seg 1/1 to: AA44485 from: 1 to: 206

57 GTCTTACCAAAATAAGCTACCATCACCATCACCAT 19
 ||||| ||||| ::|||::: ||||| ||||| |||||
 192 ValLeuProArgSerValGlyHisHisHisHisHis 204

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:AA95055

seq_documentation_block:

ID AA95055 standard; Protein; 421 AA.

XX AC AA95055;

XX 19-AUG-1996 (first entry)

DE IL-2-DETA-DGAL4 multidomain protein.

KW Nucleic acid transfer system; gene transfer; gene therapy;
 KW cell targeting; multidomain protein; vector; cancer;
 KW exotoxin A; DETA; ompA; signal peptide; GAL4; interleukin-2;
 KW IL-2.
 XX

OS Chimeric synthetic;

OS Chimeric Homo sapiens;

OS Chimeric Pseudomonas aeruginosa;

OS Chimeric Saccharomyces cerevisiae.

XX Key Location/Qualifiers

FH Peptide 1..8

FT /label= FLAG_epitope

FT Peptide 9..17

FT /label= Spacer

FT Domain 18..150


```

FT      /label= IL-2
FT      /note= "amino acids 1-113 of human IL-2"
FT      151
FT      /label= Spacer
FT      152..266
FT      /label= ETA
FT      /note= "amino acids 252-366 of ETA"
FT      267
FT      /label= Spacer
FT      268..413
FT      /label= GAL4
FT      /note= "amino acids 2-147 of yeast GAL4"
FT      414..421
FT      /label= Spacer
FT      /note= "endoplasmic reticulum retention signal"
XX
PN      W09613599-A1.
XX
XX      09-MAY-1996.
XX
PF      31-OCT-1995; 95WO-EP04270.
XX
PR      01-NOV-1994; 94EP-0810627.
XX
PA      (WELS/) WELS W.
XX
PI      Fominaya J, Wels W;
XX
DR      WPI; 1996-239505/24.
DR      N-PSDB; AAT29411.
XX
PT      Nucleic acid transfer system for gene therapy, e.g. against cancer
PT      - includes toxin translocation domain to target nucleic acid to
PT      specific cell
XX
PS      Claim 7; Page 67-69; 106pp; English.
XX
CC      A multidomain protein (AAR95055) has a FLAG epitope, a portion
CC      of human interleukin-2 that acts as a ligand domain, a
CC      non-cytotoxic portion of Pseudomonas aeruginosa exotoxin A acting
CC      as a translocation domain and the DNA binding domain of yeast GAL4.
CC      It is the product of a fusion gene (AAT29411) and can be expressed
CC      in E. coli (resulting in removal of an ompA signal peptide). It is
CC      used with an effector nucleic acid that comprises e.g. a gene to be
CC      delivered to a cell and a cognate structure for the GAL4 DNA binding
CC      domain. This provides a novel means of nucleic acid transfer,
CC      suitable for gene therapy.
XX
SQ      Sequence 421 AA;

```

```

alignment_scores:
  Quality: 60.00      Length: 13
  Ratio: 5.455      Gaps: 0
  Percent Similarity: 84.615      Percent Identity: 69.231

```

```

alignment_block:
US-09-674-779-4/rev x AAR95055 ..

```

```

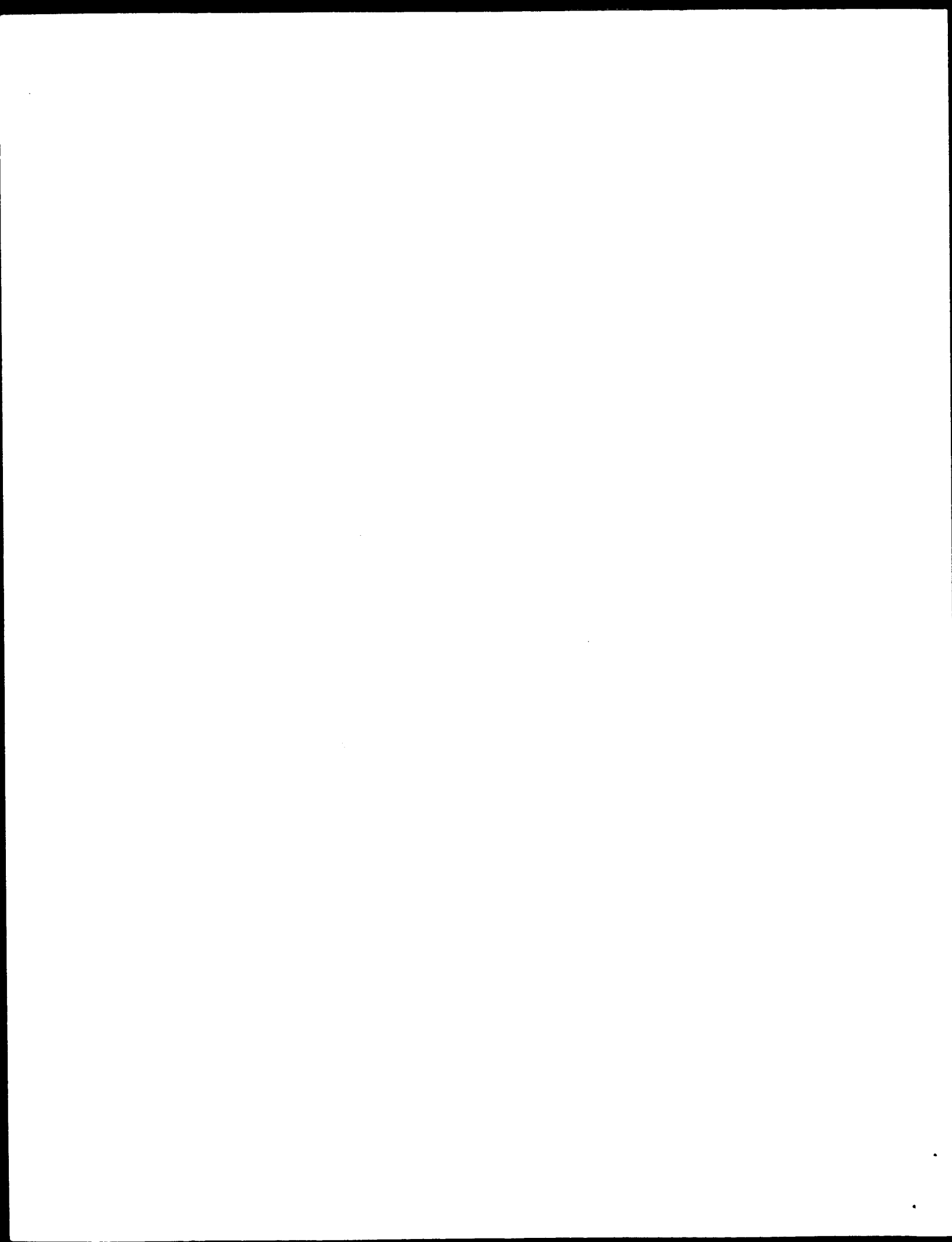
Align seg 1/1 to: AAR95055 from: 1 to: 421

```

```

45 AATAAGCTACACCATCACCATTAGTCGACCT 7
   :|||||
7 AsplysLeuHisHisHisHisHisHisLysLeuAlaPro 19

```



OM of: US-09-674-779-4 to: Issued_Patents_AA.* out_format : pfs
 Date: Jul 30, 2002 4:25 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet.n2p.model -DEV=xlp
 -O=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151752_7276/app_query.fasta_1.116
 -DB=Issued_Patents_AA -QFMR=fastan -SUFFIX=ra1 -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 -USR=US09674779.@CNI_1_29 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-4
 Query length: 60
 Database: Issued_Patents_AA.*
 Database sequences: 231628
 Database length: 2442594
 Search time (sec): 22.710000

score_list:

| Sequence | Strd Orig | ZScore | Escore | Len | Documentation |
|---|-----------|--------|--------|------|---------------|
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-447-430A-43 | 64.50 | 172.70 | 0.0116 | 338 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-244-951A-10 | 64.50 | 171.43 | 0.0118 | 391 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-389-011-23 | 64.50 | 171.43 | 0.0118 | 391 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-403-917A-23 | 64.50 | 171.43 | 0.0118 | 391 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-348-952A-23 | 64.50 | 171.43 | 0.0118 | 391 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-849-634B-3 | 63.50 | 174.59 | 0.0158 | 195 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-101-135-8 | 61.50 | 165.91 | 0.0346 | 272 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-600-783-13 | 58.00 | 179.91 | 0.0919 | 17 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-677-862-2 | 56.50 | 142.47 | 0.2491 | 763 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-252-571-2 | 56.50 | 142.47 | 0.2491 | 763 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-434-065-2 | 56.50 | 142.47 | 0.2491 | 763 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-789-275-4 | 56.50 | 142.47 | 0.2491 | 763 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-789-275-5 | 56.50 | 142.47 | 0.2491 | 763 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-524-757-30 | 56.00 | 178.74 | 0.1814 | 10 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-524-757-31 | 56.00 | 178.74 | 0.1814 | 10 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-459-064B-35 | 54.00 | 171.36 | 0.3892 | 12 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-460-421A-35 | 54.00 | 171.36 | 0.3892 | 12 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-340-812-13 | 54.00 | 153.93 | 0.4910 | 89 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-459-064B-13 | 54.00 | 153.93 | 0.4910 | 89 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-460-421A-13 | 54.00 | 153.93 | 0.4910 | 89 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-07-882-292-2 | 54.00 | 153.93 | 0.4910 | 89 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-331-644-2 | 54.00 | 139.27 | 0.5969 | 480 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-331-644-2 | 54.00 | 139.27 | 0.5969 | 480 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-041-886-21 | 54.00 | 131.43 | 0.6626 | 1182 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-149-097D-36 | 54.00 | 125.77 | 0.7144 | 2265 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-149-097D-35 | 54.00 | 124.88 | 0.7229 | 2509 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-484-956-91 | 53.00 | 140.04 | 0.8238 | 315 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-757-653-91 | 53.00 | 140.04 | 0.8238 | 315 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-757-653-163 | 53.00 | 139.90 | 0.8253 | 320 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-823-516-61 | 53.00 | 139.90 | 0.8253 | 320 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-759-038-102 | 53.00 | 139.90 | 0.8253 | 320 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-758-314-102 | 53.00 | 139.90 | 0.8253 | 320 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-572-191-4 | 53.00 | 138.52 | 0.8406 | 375 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-753-247-18 | 53.00 | 138.16 | 0.8447 | 391 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-056-556-214 | 53.00 | 134.57 | 0.8861 | 591 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-632-947B-1 | 52.50 | 138.30 | 0.9180 | 802 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-226-264-7 | 52.00 | 164.24 | 0.8323 | 14 | ! |
| /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-226-264-19 | 52.00 | 164.24 | 0.8323 | 14 | ! |

```

/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-226-264-20 - 52.00 164.24 0.8323
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-226-264-8 - 52.00 163.08 0.8453
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-226-264-21 - 52.00 163.08 0.8453
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-895-707-21 - 52.00 158.85 0.8942
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-664-449-39 - 52.00 154.68 0.9453

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-447-430A-43
seq_documentation_block:
; Sequence 43, Application US/08447430A
; Patent No. 5916558
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Recombinant polypeptides and peptides,
; TITLE OF INVENTION: nucleic acids coding for the same and use of these
; TITLE OF INVENTION: polypeptides and peptides in the diagnostic of
; NUMBER OF SEQUENCES: 43
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,430A
; FILING DATE:
; CLASSIFICATION: 424
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-08-447-430A-43

alignment_scores:
Quality: 64.50 Length: 20
Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x US-08-447-430A-43 ..
Align seg 1/1 to: US-08-447-430A-43 from: 1 to: 338
60 CATGCTTACCAAAATAAGTACCATCACCACCATTAAGTGA 11
||| : : : : : : : : : : : : : : : : : : : :
21 HisGlnValcGluGlnGlyleHisHisHisHisHisHis 36
10 CCCTCTGCCT 1
||| |||
36 pProGlyPro 39
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-244-951A-10
seq_documentation_block:
; Sequence 10, Application US/08244951A
; Patent No. 5843779
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDEMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK

```

```

; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,951A
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/03499
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP/92/403403.6
; FILING DATE: 14-DEC-1992
; APPLICATION NUMBER: 19,683
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.003A
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; FEATURE:
; NAME/KEY: mTHFMPH-taul fusion protein
; US-08-244-951A-10

alignment_scores:
  Quality: 64.50 Length: 20
  Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x US-08-244-951A-10 from: 1 to: 391
Align seg 1/1 to: US-08-244-951A-10 from: 1 to: 391
60 CATGCTTACCAAAATAAGCTACCATCACCATCACCATTAAAGTCGA 11
   |||   ::   :::::  ::|||   |||   |||   |||   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36

10 CCCTCTGCCT 1
   |||   |||
36 pProGlyPro 39

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-389-011-23

seq_documentation_block:
; Sequence 23, Application US/08389011
; Patent No. 5861257
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDEMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
;

; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,011
; FILING DATE: 15-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,917
; FILING DATE: 19-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/403,916
; FILING DATE: 19-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/244,951
; FILING DATE: 13-JUN-1994
; APPLICATION NUMBER: PCT/EP93/03499
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP/92/403403.6
; FILING DATE: 14-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CHARLES A. MUSERLIAN
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 410.003-1-CON
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391
; TYPE: Amino Acid
; STRANDEDNESS: Unknown
; TOPOLOGY: Unknown
; FEATURE:
; NAME/KEY: mTHFMPH-taul fusion protein
; US-08-244-951A-10

alignment_scores:
  Quality: 64.50 Length: 20
  Ratio: 4.300 Gaps: 1
Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:
US-09-674-779-4/rev x US-08-389-011-23 from: 1 to: 391
Align seg 1/1 to: US-08-389-011-23 from: 1 to: 391
60 CATGCTTACCAAAATAAGCTACCATCACCATCACCATTAAAGTCGA 11
   |||   ::   :::::  ::|||   |||   |||   |||   |||
21 HisGlnValGluGluGlnGlyIleHisHisHisHisHis...Valas 36

10 CCCTCTGCCT 1
   |||   |||
36 pProGlyPro 39

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.US-08-403-917A-23

seq_documentation_block:
; Sequence 23, Application US/08403917A
; Patent No. 6010913
; GENERAL INFORMATION:
; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
; APPLICANT: VANDEMEEREN, EUGEN; VAN DE VOORDE, ANDRE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: DIRECTED AGAINST THE MICROTUBULE-ASSOCIATED
; TITLE OF INVENTION: PROTEIN TAU, HYBRIDOMAS SECRETING THESE
; TITLE OF INVENTION: ANTIBODIES, ANTIGEN RECOGNITION BY THESE
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AND THEIR APPLICATIONS.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN & MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
;

```

; GENERAL INFORMATION:
 ; APPLICANT: VANDERMEEREN, MARC; MERCKEN, MARC;
 ; APPLICANT: VANMECHELEN, EUGEN;
 ; APPLICANT: VAN DE VOORDE, ANDRE
 ; TITLE OF INVENTION: MONOCLONAL ANTIBODIES

```
seq_name: /cgn_6/ptodata/2/iaa/6B_COMB.pep:US-08-849-634B-3
seq_documentation_block:
; Sequence 3, application US/08849634B
; Patent No. 6296855
```

```
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 2551-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-101-135-8
alignment_scores:
    Quality: 61.50      Length: 18
    Ratio: 4.393        Gaps: 1
    Percent Similarity: 77.778   Percent Identity: 55.556
alignment_block:
US-09-674-779-4/rev x US-09-101-135-8 ..
Align seg 1/1 to: US-09-101-135-8 from: 1 to: 272
60 CATGCTTTACCAAAATAAGCTACACCATCACCATCACCATTAAAGTCGA 11
||| ::: :::: ::::: ||||||||||||||| |||||
21 HisGlnValGluGlnGlyIleHisHisHisHisHis...ValAs 36
10 CCCT 7
|||
36 pPro 37
seq_name: /cgn_2/ptodata/2/iaa/5B_COMB pep:US-08-600-783-13
seq_documentation_block:
; Sequence 13, Application US/08600783
; Patent No. 5962267
; GENERAL INFORMATION:
; APPLICANT: SHIN, Hang Cheol
; APPLICANT: CHANG, Seung Gu
; APPLICANT: KIM, Dae Young
; APPLICANT: KIM, Chong Suhll
; TITLE OF INVENTION: Proinsulin Derivative and Process
; TITLE OF INVENTION: For Producing Human Insulin
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHIN, Hang Cheol
; STREET: Ssangma-Hanshin Apt. 102-1206,
; STREET: #245 Cholsan-dong
; CITY: Kwangmyung-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 423-030
; ADDRESSEE: CHANG, Seung Gu
; STREET: Hyundai Apt. 71-203, Apkujong-dong,
; STREET: Kangnam-ku
; CITY: Seoul
; STATE: Seoul
; COUNTRY: Republic of Korea
; ZIP: 135-110
; ADDRESSEE: KIM, Dae Young
; STREET: Sosa Jukong Apt. 108-202, Sosa Bon-dong,
; STREET: Sosa-ku
; CITY: Bucheon-shi
; STATE: Kyungki-do
; COUNTRY: Republic of Korea
; ZIP: 422-230
; ADDRESSEE: KIM, Chong Suhll
; STREET: Garden Heights Apt. 202-801, #100,
; STREET: Hwangkeum-dong, Soosung-ku
; CITY: Taegu
; STATE: Taegu
```

```

; COUNTRY: Republic of Korea
; ZIP: 706-040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk, 3.5 inch, 1.44MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,783
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: KR 95-2751
; FILING DATE: 15-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Shahan Islam
; REGISTRATION NUMBER: 32,507
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-1000
; TELEFAX: (212) 953-7249
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-600-783-13

alignment_scores:
  Quality: 58.00      Length: 11
  Ratio: 5.800       Gaps: 0
Percent Similarity: 90.909 Percent Identity: 72.727

alignment_block:
US-09-674-779-4/rev x US-08-600-783-13 ..
Align seg 1/1 to: US-08-600-783-13 from: 1 to: 17

51 CCAAAATAAGCTACACATCACCACCAT 19
|||||
1 ProSerAspLysProHisHisHisHis 11

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-677-862-2
seq_documentation_block:
; Sequence 2, Application US/08677862
; Patent No. 5874230
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,862
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J

```

```

; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-677-862-2

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708       Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
US-09-674-779-4/rev x US-08-677-862-2 ..
Align seg 1/1 to: US-08-677-862-2 from: 1 to: 763

60 CATGCTTACAAA.....AATAAGCT 38
|||||
590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 606
|||||
37 ACACCATCACCACCATCAT 19
|||||
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-09-252-571-2
seq_documentation_block:
; Sequence 2, Application US/09252571
; Patent No. 5981250
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/252,571
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-252-571-2

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
US-09-674-779-4/rev x US-09-252-571-2
Align seg 1/1 to: US-09-252-571-2 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-434-065-2

seq_documentation_block:
; Sequence 2, Application US/09434065
; Patent No. 6107074
; GENERAL INFORMATION:
; APPLICANT: SONG, Ho Yeong
; APPLICANT: ROTHE, Mike
; TITLE OF INVENTION: TRAF2-Associated Kinase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, 34th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/434,065
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/677,862
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brezner, David J
; REGISTRATION NUMBER: 24,774
; REFERENCE/DOCKET NUMBER: T96-005/A63613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415 781-1989
; TELEFAX: 415 398-3249
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-434-065-2

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1

```

```

Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
US-09-674-779-4/rev x US-09-434-065-2
Align seg 1/1 to: US-09-434-065-2 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-789-275-4

seq_documentation_block:
; Sequence 4, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome
; FILE REFERENCE: U 011114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 4
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; US-08-789-275-4

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
US-09-674-779-4/rev x US-08-789-275-4
Align seg 1/1 to: US-08-789-275-4 from: 1 to: 763

60 CATGCTTTACCAAAA.....AATAAGCT 38
||||| |||:::
590 HisValAlaProGlnGlnAsnAlaLeuHisHisGlyAsnSerSe 606
37 ACACCATCACCATCACCAT 19
||||| |||:::
606 rHisHisHisHisHisHis 612

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-789-275-5

seq_documentation_block:
; Sequence 5, Application US/08789275A
; Patent No. 6251664
; GENERAL INFORMATION:
; APPLICANT: Palreja, Estivill Xavier
; APPLICANT: Pritchard, Melanie
; APPLICANT: Vilario, Jordi Guimera
; TITLE OF INVENTION: Human Gene Sequence of the Down Syndrome Critical
; Patent No. 6251664
; TITLE OF INVENTION: Region of Human Chromosome 21, Coding for A Serine-
; TITLE OF INVENTION: Threonine Protein Kinase (MNB), Expressed in the
; TITLE OF INVENTION: Neuronal Regions Affected in Down Syndrome

```



```

; FILE REFERENCE: U 011114-4
; CURRENT APPLICATION NUMBER: US/08/789,275A
; CURRENT FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 5
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Rat norvegicus
; FEATURE:
US-08-789-275-5

```

```

alignment_scores:
  Quality: 56.50      Length: 23
  Ratio: 4.708        Gaps: 1
  Percent Similarity: 52.174  Percent Identity: 43.478

```

```

alignment_block:
US-09-674-779-4/rev x US-08-789-275-5 ..

```

```

Align seg 1/1 to: US-08-789-275-5 from: 1 to: 763

```

```

60 CATGCTTACCAAAA.....ATAACCT 38
||||| ||||| |||||
590 HisValAlaProGlnAsnAlaLeuHisHisHisGlyAsnSer 606
37 ACACCATCACCATCACCAT 19
||||| ||||| |||||
606 rHisHisHisHisHisHis 612

```

```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-524-757-30

```

```

seq_documentation_block:
; Sequence 30, Application US/08524757
; Patent No. 5792634
; GENERAL INFORMATION:
; APPLICANT: Conaway, Ronald C.
; APPLICANT: Conaway, Joan W.
; APPLICANT: Bradsher, John N.
; TITLE OF INVENTION: RNA Polymerase Transcription Factor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13621
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160087
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Harre, John A.
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35006CIPCIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (214) 939-4500
; TELEFAX: (214) 939-4600
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-524-757-30

```

```

alignment_scores:
  Quality: 56.00      Length: 10
  Ratio: 6.222        Gaps: 0
  Percent Similarity: 90.000  Percent Identity: 80.000

```

```

alignment_block:
US-09-674-779-4/rev x US-08-524-757-30 ..

```

```

Align seg 1/1 to: US-08-524-757-30 from: 1 to: 10

```

```

39 CTACACCATCACCATCACCATTAAAGTCGAC 10
:::||||| ||||| |||||
1 MetHisHisHisHisHisHisAsnValAsp 10

```

```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-524-757-31

```

```

seq_documentation_block:
; Sequence 31, Application US/08524757
; Patent No. 5792634
; GENERAL INFORMATION:
; APPLICANT: Conaway, Ronald C.
; APPLICANT: Conaway, Joan W.
; APPLICANT: Bradsher, John N.
; TITLE OF INVENTION: RNA Polymerase Transcription Factor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13621
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160087
; FILING DATE: 30-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Harre, John A.
; REGISTRATION NUMBER: 37,345
; REFERENCE/DOCKET NUMBER: B35006CIPCIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (214) 939-4500
; TELEFAX: (214) 939-4600
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-524-757-31

```

```

alignment_scores:
  Quality: 56.00      Length: 10

```

Ratio: 6.222 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x US-08-524-757-31 ..

Align seg 1/1 to: US-08-524-757-31 from: 1 to: 10

39 CTACACCATCACCATCACCATTAAAGTCGAC 10
:::|||||
1 MethHisHisHisHisHisValAsp 10

OM of: US-09-674-779-4 to: PIR_71.* out_format : pfs

Date: Jul 30, 2002 4:25 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame2_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151753_7289/app_query.fasta_1.116
-DB=PIR_71 -OPMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINARCH=0.100 -LOOFCU=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-QGAPEXT=0.050 -YGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEADSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USBR=US09674779_@CGNI_1.113
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIME=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: PIR_71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 28.830000
```

score_list:

| Sequence | Strd Orig | Zscore | Escore | Len | Documentation |
|-------------|-----------|--------|--------|------|-----------------------------------|
| pir2:T04874 | 64.00 | 164.71 | 0.0416 | 1057 | ! hypothetical protein F28A21.230 |
| pir2:S60988 | 62.50 | 167.77 | 0.0726 | 409 | ! hypothetical protein YOR134W - |
| pir1:OMRTSP | 61.00 | 163.84 | 0.1276 | 385 | ! selenoprotein P precursor [vali |
| pir2:A1907 | 59.00 | 156.31 | 0.2713 | 476 | ! methyl CpG binding protein 2 - |
| pir2:T31611 | 59.00 | 156.05 | 0.2714 | 492 | ! methyl-CpG-binding protein 2 - |
| pir2:A56235 | 58.00 | 143.93 | 0.3989 | 1585 | ! hypothetical protein Y50F8A.g |
| pir2:T46375 | 57.00 | 153.79 | 0.5740 | 311 | ! transcription activator WafB - |
| pir2:T45727 | 57.00 | 151.08 | 0.5755 | 439 | ! hypothetical protein DKF2p434D1 |
| pir2:JC4898 | 56.50 | 154.36 | 0.6915 | 240 | ! hypothetical protein FIP2.220 - |
| pir2:T32443 | 56.00 | 145.36 | 0.6975 | 754 | ! Down-syndrome-critical-region B |
| pir1:A38565 | 56.00 | 153.61 | 0.8341 | 219 | ! hypothetical protein T28B4.4 - |
| pir2:T37451 | 56.00 | 149.07 | 0.8377 | 390 | ! polycarb (PC) protein - fruit f |
| pir2:A54743 | 56.00 | 147.63 | 0.8389 | 469 | ! HGF-G2 (HFG-2) protein - human |
| pir2:A43817 | 56.00 | 147.51 | 0.8390 | 476 | ! transcription factor HFK1 - hum |
| pir2:S27473 | 56.00 | 142.45 | 0.8431 | 906 | ! transforming protein (cbl) - hu |
| pir2:T23056 | 56.00 | 142.08 | 0.8434 | 950 | ! URBS1 protein - smut fungus (Us |
| pir2:C96478 | 56.00 | 138.68 | 0.8462 | 1465 | ! hypothetical protein H06001.2 |
| pir2:T47447 | 55.50 | 136.69 | 0.8478 | 1887 | ! protein F1504.13 [imported] - |
| pir2:T24446 | 55.00 | 133.58 | 1.02 | 2326 | ! calcium channel protein alpha - |
| pir2:S23737 | 55.00 | 151.08 | 1.21 | 208 | ! hypothetical protein T04C10.4 - |
| pir2:T49529 | 55.00 | 148.28 | 1.22 | 297 | ! proline-rich protein precursor |
| pir2:S39406 | 55.00 | 147.62 | 1.22 | 323 | ! transcription factor-kr - mouse |
| pir2:S95345 | 55.00 | 146.90 | 1.22 | 354 | ! homeotic protein otx1 - human |
| pir2:T56547 | 55.00 | 146.88 | 1.22 | 355 | ! otx1 protein - mouse |
| pir2:A49077 | 55.00 | 146.88 | 1.22 | 355 | ! homeodomain protein otx1 - rat |
| pir1:A56018 | 55.00 | 146.43 | 1.22 | 378 | ! transcription initiation factor |
| pir1:S30205 | 55.00 | 145.05 | 1.22 | 448 | ! transcription factor Oct-6 - hu |
| pir1:A40168 | 55.00 | 145.03 | 1.22 | 449 | ! transcription factor Oct-6 - mc |
| pir2:T33997 | 55.00 | 145.00 | 1.22 | 451 | ! transcription factor Oct-6 - ra |
| pir1:A40879 | 55.00 | 144.66 | 1.22 | 471 | ! hypothetical protein W03G1.5 - |
| pir1:A40879 | 55.00 | 136.65 | 1.23 | 1305 | ! phospholipase C (EC 3.1.4.3), |
| pir2:T46477 | 55.00 | 136.61 | 1.23 | 1312 | ! phospholipase C (EC 3.1.4.3), |
| pir2:T47773 | 55.00 | 132.29 | 1.24 | 2273 | ! calcium channel BI-1 - rabbit |
| pir2:AA0721 | 54.00 | 141.79 | 1.78 | 467 | ! hypothetical protein F2A13.210 |
| pir2:JH0672 | 54.00 | 141.67 | 1.78 | 474 | ! neuroblast proliferation inhibi |
| pir1:S31223 | 54.00 | 141.58 | 1.78 | 480 | ! brain factor I protein - rat |
| pir2:T17291 | 54.00 | 141.33 | 1.78 | 495 | ! transcription factor Brn-1 - mc |
| pir2:T13958 | 54.00 | 137.13 | 1.79 | 845 | ! hypothetical protein DKF2p434E1 |
| pir2:T13804 | 54.00 | 134.60 | 1.79 | 1166 | ! synGAP-bl protein - rat |
| | 54.00 | 134.30 | 1.79 | 1212 | ! shs protein - fruit fly (Dros |

```
pir2:TL4270 - 54.00 134.06 1.79 1249 ! Ras-GTPase activating prote
pir2:TL4259 - 54.00 133.79 1.79 1293 ! ras GTPase-activating prote
pir2:T06387 - 53.50 145.20 2.14 251 ! knotted I class homeodomain
pir2:A41098 - 53.50 128.11 2.17 221 ! calcium channel protein alp
pir2:S37939 - 53.00 150.15 2.57 111 ! hypothetical protein YKL111c
```

seq_name: pir2:T04874

seq_documentation_block:

```
hypothetical protein F28A21.230 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T04874
R:Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15387
A:Accession: T04874
A:Molecule type: DNA
A:Residues: 1-1057 <BEV>
A:Cross-references: EMBL:AL035526
A:Experimental source: cultivar Columbia; BAC clone F28A21
C:Genetics:
A:Map position: 4
A:Introns: 680/3: 754/1; 881/1
A:Note: F28A21.230
```

alignment_scores:

```
Quality: 64.00 Length: 11
Ratio: 5.818 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 72.727
```

alignment_block:

```
US-09-674-779-4/rev x T04874 ..
Align seg 1/1 to: T04874 from: 1 to: 1057
51 CCAAAATAAGCTACCATCATCATCATCAT 19
||||:||||:||||:||||:||||:||||:
928 ProARGASnVnValHishHisHisHis 938
```

seq_name: pir2:S60988

seq_documentation_block:

```
hypothetical protein YOR134W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O3320; hypothetical protein YOR3320W
C:Species: Saccharomyces cerevisiae
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 07-May-1999
C:Accession: S60988; S61690; S67019; S63865
R:Wiemann, S.; Rechmann, S.; Benes, V.; Voss, H.; Schwager, C.; Vleck, C.; Stegemann,
submitted to the EMBL Data Library, August 1995
A:Description: Sequencing of 51 kilobases on the right arm of chromosome XV from S. c
A:Reference number: S60983
A:Accession: S60988
A:Molecule type: DNA
A:Residues: 1-409 <WIE>
A:Cross-references: EMBL:X90518; NID:g1050808; PID:g1050814
R:Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valen
submitted to the EMBL Data Library, December 1995
A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromos
A:Reference number: S61643
A:Accession: S61690
A:Molecule type: DNA
A:Residues: 1-409 <BEN>
A:Cross-references: EMBL:X94335; NID:g1262139; PID:e217752; PID:g1164978
R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansorge, W
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66965
A:Accession: S67019
A:Molecule type: DNA
A:Residues: 1-409 <VOS>
A:Cross-references: EMBL:Z75042; NID:g1420343; PID:e252026; PID:g1420344; MIPS:YOR134
A:Experimental source: strain S288C
```

```

alignment_block:
  US-09-674-779-4/rev x OMR TSP      ..
  Align seg 1/1 to: OMR TSP from: 1 to: 385
  54 TTACCAAAAAATAAGCTACCATCACCATCACCATT 19
  ||||| :: |||||||||||||||||
  238 LeuProSerGlyLeuHisHisHisHisHisHis 249

seq_name: pir2:S57963

seq_documentation_block:
  methyl CpG binding protein 2 - human (fragment)
  C:Species: Homo sapiens (man)
  C:CDate: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Nov-1999
  C:Accession: S57963
  A:Submitted to: R.R.d'Esposito, M.; Quaderli, N.A.; Ciccodicola, A.; Bruni, P.; Esposito, T.; D'Urso, M.
  A:Description: Physical mapping and expression analysis of an X-linked gene encoding
  A:Reference number: S57963
  A:Accession: S57963
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-476 <DES>
  A:Cross-references: EMBL, X89430; NID:G899295; PIDN:CAAG1599.1; PID:q899296

```

60 CATCTCTTACCAAAAAATAGCA...
3302 HisMetLeuProArgSerArgAlaLeuSerAspSerAsnAsnPhethrI 318

36CACCATCACCATTAACTGCGACCTTCGCCT 1
378 ewishshishishishishishisAlaLeuPheProSerPro 332

seq_documentation_block:
 C:Species: Rattus norvegicus [validated] - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 15-Sep-2000
 C:Accession: A40380; B40380; S68322
 C:Accession: A40380; B40380; S68322
 J:Bill, K.E.; Lloyd, R.S.; Yang, J.G.; Read, R.; Burk, R.F.
 J:Hiol. Chem. 266, 10050-10053, 1991
 A:title: The cDNA for rat selenoprotein P contains 10 TGA codons in the open
 A:Reference number: A40380; MUID:91244760
 A:Accession: A40380
 A:Molecule type: mRNA
 A:Residues: 1-385 <HI>
 A:Cross-references: GB:M63574; NID:g206893; PIDN:AAA42129.1; PID:g206894
 A:Accession: B40380
 A:Molecule type: protein
 A:Residues: 20-41; 267-287; 316-327 <HI2>
 R:Chittum, H.S.; Himeeno, S.; Hill, K.E.; Burk, R.F.
 Arch. Biochem. Biophys. 325, 124-128, 1996
 A:title: Multiple forms of selenoprotein P in rat plasma.
 A:Reference number: S68322; MUID:96140605
 A:Accession: S68322
 A:Molecule type: protein
 A:Residues: 20-27 <CHI>

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 61.00 | Length: 12 |
| Ratio: | 6.100 | Gaps: 0 |
| Percent Similarity: | 83.333 | percent Identity: 75.000 |

```
alignment_scores:
  Quality: 59.00      Length: 15
  Ratio: 4.917       Gaps: 0
  Percent Similarity: 80.000    Percent Identity: 60.000

alignment_block:
```


Align seg 1/1 to: T45727 from: 1 to: 240

48 AAAAATAAGTACACCATCACCATTAAGTCGACCTCTGCCT 1
:::|||||
17 GluAsnGlnIleHisHisHisHis.....SerProLeupro 29

seq_name: pir2:JC4898

seq_documentation_block:

Down-syndrome-critical-region protein - human
N:Alternate names: Drosophila minibrain protein homolog
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JC4898
R:Shindoh, N.; Kudoh, J.; Maeda, H.; Yamaki, A.; Minoshima, S.; Shimizu, Y.; Shimizu, N.
Biochem. Biophys. Res. Commun. 225, 92-99, 1996
A:Title: Cloning of a human homolog of the Drosophila minibrain/rat dyrk gene from "the
A:Reference number: JC4898; MUID:96332410
A:Accession: JC4898
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-754 <SHI>
A:Cross-references: DDBJ:D85759; NID:g1526445; PIDN:BAAL2866.1; PID:g1526446
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

alignment_scores:
Quality: 56.50 Length: 23
Ratio: 4.708 Gaps: 1
Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
US-09-674-779-4/rev x JC4898 ..

Align seg 1/1 to: JC4898 from: 1 to: 754

60 CATGCTTACCAAAA.....AATAAGCT 38
|||||
581 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSerSe 597
37 ACACCATCACCATCACCAT 19
|||||
597 rHisHisHisHisHisHis 603

seq_name: pir2:T32443

seq_documentation_block:

hypothetical protein T28B4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T32443
R:Wilson, R.; Greco, T.; Sansone, J.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid T28B4.
A:Reference number: Z21168
A:Accession: T32443
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-219 <WIL>
A:Cross-references: EMBL:AF026206; PIDN:AAB71260.1; GSPDB:GN00028; CESP:T28B4.4
A:Experimental source: strain Bristol N2; clone T28B4
C:Genetics:
A:Gene: CESP:T28B4.4
A:Map position: X
A:Introns: 34/2; 138/2; 184/3

alignment_scores:
Quality: 56.00 Length: 9
Ratio: 6.222 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 77.778

alignment_block:

US-09-674-779-4/rev x T32443 ..

Align seg 1/1 to: T32443 from: 1 to: 219

45 AATAAGCTACACCATCACCATCACCAT 19
:::|||||
35 HisLysMetHisHisHisHisHis 43

seq_name: pir1:A38565

seq_documentation_block:

polycarb (Pc) protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A38565
R:Paro, R.; Hogness, D.S.
Proc. Natl. Acad. Sci. U.S.A. 88, 263-267, 1991
A:Title: The polycarb protein shares a homologous domain with a heterochromatin-associ
A:Reference number: A38565; MUID:91095442
A:Accession: A38565
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-390 <PAR>
A:Cross-references: GB:X55702; NID:g8321; PIDN:CAA39229.1; PID:g603986
C:Genetics:
A:Gene: FlyBase:Pc
A:Cross-references: FlyBase:FBgn0003042
C:Superfamily: polycarb protein; chromobox homolog
F:26-63/Domain: chromobox homolog <CBH>

alignment_scores:
Quality: 56.00 Length: 10
Ratio: 6.222 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x A38565 ..

Align seg 1/1 to: A38565 from: 1 to: 390

48 AAAATAAGCTACACCATCACCATCACCAT 19
|||||
130 LysGluLysLysHisHisHisHisHisHis 139

seq_name: pir2:I37451

seq_documentation_block:

HBF-G2 (HFK-2) protein - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 24-Sep-1999
C:Accession: I37451
R:Wiese, S.; Murphy, D.B.; Schlung, A.; Burfeind, P.; Schmundt, D.; Schnulle, V.; Mat
Biochim. Biophys. Acta 1262, 105-112, 1995
A:Title: The genes for human brain factor 1 and 2, members of the fork head gene fami
A:Reference number: I37451; MUID:95322450
A:Accession: I37451
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-469 <RES>
A:Cross-references: EMBL:X78202; NID:g967047; PIDN:CAA55038.1; PID:g967048
C:Genetics:
A:Gene: HBF-G2; HFK-2
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homolog
F:162-253/Domain: fork head DNA-binding domain homolog <FHD>

alignment_scores:
Quality: 56.00 Length: 17
Ratio: 5.091 Gaps: 0
Percent Similarity: 64.706 Percent Identity: 52.941

OM of: US-09-674-779-4 to: SwissProt_40:* out_format : pfs

Date: Jul 30, 2002 4:31 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (C) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09674779/runat_30072002_151754_7376/app_query.fasta_1.116
-DB=SwissProt_40 -OPMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_CGNI_1.45 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-4

Query length: 60

Database: SwissProt_40:*

Database sequences: 105224

Database length: 38719550

Search time (sec): 16.440000

score_list:

| Sequence | Strd | Orig | Zscore | Escore | Len | Documentation |
|-------------------------|-------|--------|--------|--------|-----|--------------------------------|
| SwissProt_40:SELP_RAT | 61.00 | 159.40 | 0.0909 | 385 | 1 | P25236 rattus norvegicus (rat) |
| SwissProt_40:MBC2_HUMAN | 59.00 | 151.91 | 0.1882 | 486 | 1 | P51608 homo sapiens (human) |
| SwissProt_40:MBC2_RAT | 59.00 | 151.81 | 0.1882 | 492 | 1 | Q00568 rattus norvegicus (rat) |
| SwissProt_40:MBC2_MOUSE | 58.00 | 149.12 | 0.2704 | 484 | 1 | Q922d6 mus musculus (mouse) |
| SwissProt_40:SNF1_CANGA | 57.00 | 144.45 | 0.3896 | 611 | 1 | Q00372 candida glabrata (yea) |
| SwissProt_40:DYRA_HUMAN | 56.50 | 141.29 | 0.4682 | 763 | 1 | Q13627 homo sapiens (human) |
| SwissProt_40:DYRA_MOUSE | 56.50 | 141.29 | 0.4682 | 763 | 1 | Q61214 mus musculus (mouse) |
| SwissProt_40:DYRA_RAT | 56.50 | 141.29 | 0.4682 | 763 | 1 | Q63470 rattus norvegicus (rat) |
| SwissProt_40:PC_DROME | 56.00 | 145.17 | 0.5567 | 390 | 1 | P26017 drosophila melanogaster |
| SwissProt_40:FXGA_HUMAN | 56.00 | 143.72 | 0.5580 | 469 | 1 | P55316 homo sapiens (human) |
| SwissProt_40:FXGB_HUMAN | 56.00 | 143.58 | 0.5581 | 477 | 1 | P55315 homo sapiens (human) |
| SwissProt_40:GEL_HUMAN | 56.00 | 138.52 | 0.5624 | 906 | 1 | P22681 homo sapiens (human) |
| SwissProt_40:URB1_USTMA | 56.00 | 138.15 | 0.5627 | 950 | 1 | P40349 ustilago maydis (smut) |
| SwissProt_40:CCAB_DIOSM | 55.50 | 129.67 | 0.6818 | 2326 | 1 | P56698 discopyge ommata (el) |
| SwissProt_40:MAF1_MOUSE | 55.00 | 143.83 | 0.7981 | 323 | 1 | P54841 mus musculus (mouse) |
| SwissProt_40:MAF1_RAT | 55.00 | 143.83 | 0.7981 | 323 | 1 | P54842 rattus norvegicus (rat) |
| SwissProt_40:OTX1_HUMAN | 55.00 | 143.11 | 0.7989 | 354 | 1 | P32242 homo sapiens (human) |
| SwissProt_40:OTX1_MOUSE | 55.00 | 143.09 | 0.7990 | 355 | 1 | P80205 mus musculus (mouse) |
| SwissProt_40:OTX1_RAT | 55.00 | 143.09 | 0.7990 | 355 | 1 | P80205 mus musculus (mouse) |
| SwissProt_40:22AA_HUMAN | 55.00 | 142.63 | 0.7995 | 355 | 1 | Q63410 rattus norvegicus (rat) |
| SwissProt_40:OCT6_HUMAN | 55.00 | 141.25 | 0.8012 | 448 | 1 | P52655 homo sapiens (human) |
| SwissProt_40:OCT6_MOUSE | 55.00 | 141.23 | 0.8012 | 448 | 1 | Q03052 homo sapiens (human) |
| SwissProt_40:OCT6_RAT | 55.00 | 141.20 | 0.8013 | 451 | 1 | P21952 mus musculus (mouse) |
| SwissProt_40:PIPL_DROME | 55.00 | 132.77 | 0.8116 | 1312 | 1 | P20267 rattus norvegicus (rat) |
| SwissProt_40:C258_HUMAN | 54.00 | 143.38 | 1.114 | 239 | 1 | P54545 drosophila melanogaster |
| SwissProt_40:ANA_DROME | 54.00 | 137.98 | 1.15 | 474 | 1 | P58505 homo sapiens (human) |
| SwissProt_40:FXGB_RAT | 54.00 | 137.88 | 1.15 | 480 | 1 | Q26307 drosophila melanogaster |
| SwissProt_40:FXGB_MOUSE | 54.00 | 137.87 | 1.15 | 481 | 1 | Q00939 rattus norvegicus (rat) |
| SwissProt_40:ONC2_HUMAN | 54.00 | 137.80 | 1.15 | 481 | 1 | Q06987 mus musculus (mouse) |
| SwissProt_40:BRN1_MOUSE | 54.00 | 137.64 | 1.15 | 485 | 1 | Q09548 homo sapiens (human) |
| SwissProt_40:BRN1_RAT | 54.00 | 137.61 | 1.15 | 495 | 1 | P31361 mus musculus (mouse) |
| SwissProt_40:BRN1_HUMAN | 54.00 | 137.61 | 1.15 | 497 | 1 | Q63262 rattus norvegicus (rat) |
| SwissProt_40:CBX4_HUMAN | 54.00 | 137.56 | 1.15 | 500 | 1 | P20264 homo sapiens (human) |
| SwissProt_40:ECR_LUCCU | 54.00 | 136.69 | 1.15 | 558 | 1 | P20264 homo sapiens (human) |
| SwissProt_40:PHYB_SORBI | 54.00 | 134.29 | 1.16 | 757 | 1 | O18531 lucilia cuprina (green) |
| SwissProt_40:BUN2_DROME | 54.00 | 130.80 | 1.17 | 1178 | 1 | P93527 sorghum bicolor (sorgh) |
| SwissProt_40:CCAA_HUMAN | 54.00 | 130.58 | 1.17 | 1211 | 1 | Q24523 drosophila melanogaster |
| SwissProt_40:CCAA_MOUSE | 54.00 | 124.84 | 1.18 | 2505 | 1 | O00555 homo sapiens (human) |
| SwissProt_40:CCAA_RAT | 53.50 | 124.59 | 1.41 | 2164 | 1 | P97445 mus musculus (mouse) |
| SwissProt_40:YKLI_YEAST | 53.50 | 124.41 | 1.41 | 2212 | 1 | P54282 rattus norvegicus (rat) |
| SwissProt_40:YKLI_MOUSE | 53.00 | 146.61 | 1.63 | 111 | 1 | P36074 saccharomyces cerevis |

SwissProt_40:CAR2_DICDI - 53.00 137.00 1.65 375 ! P34907 dictyostelium disc
SwissProt_40:SELP_MOUSE - 53.00 136.90 1.65 380 ! P70274 mus musculus (mous
SwissProt_40:FXB2_MOUSE - 53.00 135.96 1.65 428 ! Q64733 mus musculus (mous
SwissProt_40:PDM1_DROME - 53.00 133.28 1.66 601 ! P31368 drosophila melanog
SwissProt_40:YG51_YEAST - 53.00 131.17 1.67 785 ! P50089 saccharomyces cere

seq_name: SwissProt_40:SELP_RAT

seq_documentation_block:

ID SELP_RAT STANDARD; PRT; 385 AA.
AC P25236;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Selenoprotein P precursor (Sep).
GN SEPL1 OR SELP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC TISSUE=Liver;
RX MEDLINE=91244760; PubMed=2037562;
RA Hill K.E., Lloyd R.S., Yang J.-G., Read R., Burk R.F.;
RT "The cDNA for rat selenoprotein P contains 10 TGA codons in the open
reading frame.";
RL J. Biol. Chem. 266:10050-10053(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Kidney;
RX MEDLINE=95364621; PubMed=7637580;
RA Saitoh K., Saito N., Lee M.J., Fujii M., Kobayashi T., Sumino K.;
RT "Molecular cloning of cDNA encoding a bovine selenoprotein P-like
protein containing 12 selenocysteines and a (His-Pro) rich domain
insertion, and its regional expression.";
RL Brain Res. Mol. Brain Res. 30:301-311(1995).
RN [3]
RP REVIEW.
RX MEDLINE=95017128; PubMed=7931697;
RA Burk R.F., Hill K.E.;
RT "Selenoprotein P. A selenium-rich extracellular glycoprotein.";
RL J. Nutr. 124:1891-1897(1994).
CC !- FUNCTION: MIGHT BE RESPONSIBLE FOR SOME OF THE EXTRACELLULAR
CC THE ANTIOXIDANT DEFENSE PROPERTIES OF SELENIUM OR MIGHT BE INVOLVED IN
CC THE TRANSPORT OF SELENIUM.
CC !- SURCELLULAR LOCATION: Extracellular.
CC !- TISSUE SPECIFICITY: MADE IN THE LIVER AND SECRETED INTO THE
CC PLASMA.
CC !- MISCELLANEOUS: THE SELENOCYSTEINES ARE ALL ENCODED BY THE OPAL
CC CODON, UGA.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M63574; AAA42129.1; -;
DR EMBL; D25221; BAA04950.2; -;
DR PIR; A40380; OMRTSP.
KW Glycoprotein; Signal; Selenium; Selenocysteine; Plasma.
FT SIGNAL 1 19
FT CHAIN 20 385 SELENOPROTEIN P.
FT DOMAIN 244 252 POLY-HIS.
FT SE_CYS 59 59
FT SE_CYS 264 264
FT SE_CYS 282 282
FT SE_CYS 323 323
FT SE_CYS 335 335
FT SE_CYS 357 357

```

FT SE_CYS      371      371
FT SE_CYS      373      373
FT SE_CYS      380      380
FT SE_CYS      382      382
FT CARBOHYD    83      83      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    174     174      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    188     188      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    370     370      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    375     375      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE    385 AA; 42614 MW; 013BDF4FB741E3E8 CRC64;

alignment_scores:
  Quality:      61.00      Length:      12
  Ratio:        6.100      Gaps:      0
  Percent Similarity: 83.333      Percent Identity: 75.000

alignment_block:
US-09-674-779-4/rev x SEMP_RAT ..
Align seg 1/1 to: SEMP_RAT from: 1 to: 385
54 TTACCAAAAATAAGCTACATCACATCACATCAT 19
||||| :||| ||||| ||||| |||||
238 LeuProSerGlyLeuHisHisHisHisHis 249

seq_name: SwissProt_40:MEC2_HUMAN

seq_documentation_block:
ID MEC2_HUMAN STANDARD; PRT; 486 AA.
AC P51608; O15233;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Methyl-CpG-binding protein 2 (MeCP2 protein) (MeCP2).
GN MECP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Kudo S., Fukuda M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Thiesen J., Straetling W.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97130625; PubMed=8976388;
RA Vilain A., Apioy F., Vogt N., Dutrillaux B., Malfoy B.;
RT "Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization."
RL Cytogenet. Cell Genet. 74:293-294 (1996).
[4]
RP SEQUENCE FROM N.A.
RA Reichwald K., Rosenthal A., Kioschis P., Platzer M.;
RT "Mapping and sequence analysis of the human MeCP2 locus."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RX MEDLINE=95299240; PubMed=10369871;
RA Coy J.F., Sedlacek Z., Baechner D., Dellus H., Poustka A.;
RT "A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression."
RL Hum. Mol. Genet. 8:1253-1262 (1999).
[6]
RP SEQUENCE OF 10-486 FROM N.A.
RT TISSUE=Skeletal muscle;

```

```

RX MEDLINE=96327611; PubMed=8672133;
RA D'Esposito M., Quaderi N.A., Ciccodicola A., Bruni P., Esposito T.,
RT "Isolation, physical mapping, and Northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2."
RL Mamm. Genome 7:533-535 (1996).
[7]
RP SEQUENCE OF 10-486 FROM N.A.
RA Reichwald K., Bauer D., Brenner V., Drescher B., Coy J.,
RA Kioschis P., Korn B., Nyakatura G., Platzer M., Poustka A.,
RA Sandoval N., Rosenthal A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[8]
RP VARIANTS RTT TRP-106; CYS-133; SER-155 AND MET-158.
RX MEDLINE=99438392; PubMed=10508514;
RA Amir R.E., Van den Veyver I.B., Wan M., Tran C.Q., Francke U.,
RT Zoghbi H.Y.;
RA "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2."
RL Nat. Genet. 23:185-188 (1999).
[9]
RP VARIANT RTT VAL-140.
RX MEDLINE=20465115; PubMed=11007980;
RA Orrico A., Lam C., Galli L., Dotti M.T., Hayek G., Tong S.F.,
RA Poon P.M., Zappella M., Federico A., Sorrentino V.;
RT "MECP2 mutation in male patients with non-specific X-linked mental retardation."
RL FEBS Lett. 481:285-288 (2000).
[10]
RP VARIANTS RTT W-106; F-124; C-13; C-134; R-152; M-158 AND C-306.
RX MEDLINE=20439334; PubMed=10991688;
RA Obata K., Matsuishi T., Yamashita Y., Fukuda T., Kuwajima K.,
RA Horiuchi I., Nagamitsu S., Iwanaga R., Kimura A., Omori I., Endo S.,
RA Mori K., Kondo I.;
RT "Mutation analysis of the methyl-CpG binding protein 2 gene (MECP2) in patients with Rett syndrome."
RL J. Med. Genet. 37:608-610 (2000).
[11]
RP VARIANTS RTT R-101; W-106; M-158 AND C-306, AND VARIANT K-397.
RX MEDLINE=20439335; PubMed=10991689;
RA Hampson K., Woods C.G., Latif F., Webb T.;
RT "Mutations in the MECP2 gene in a cohort of girls with Rett syndrome."
RL J. Med. Genet. 37:610-612 (2000).
[12]
RP VARIANT RTT VAL-140.
RX PubMed=11805248;
RA Dotti M.T., Orrico A., De Stefano N., Battisti C., Sicurelli F.,
RA Severi S., Lam C.W., Galli L., Sorrentino V., Federico A.;
RT "A Rett syndrome MECP2 mutation that causes mental retardation in men."
RL Neurology 58:226-230 (2002).
CC -!- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN BIND SPECIFICALLY TO A SINGLE METHYL-CpG PAIR. IT IS NOT INFLUENCED BY SEQUENCES FLANKING THE METHYL-CpGS. MEDIATES TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE DEACETYLASE AND THE COREPRESSOR SIN3A.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CpG IN THE GENOME.
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL ADULT SOMATIC TISSUES TESTED.
CC -!- DISEASE: DEFECTS IN MECP2 ARE THE CAUSE OF RETT SYNDROME (RTT), AN X-LINKED DOMINANT DISEASE. RTT IS A PROGRESSIVE NEUROLOGIC DEVELOPMENTAL DISORDER AND ONE OF THE MOST COMMON CAUSES OF MENTAL RETARDATION IN FEMALES. PATIENTS APPEAR TO DEVELOP NORMALLY UNTIL 6 TO 18 MONTHS OF AGE, THEN GRADUALLY LOOSE SPEECH AND PURPOSEFUL HAND MOVEMENTS AND DEVELOP MICROCEPHALY, SEIZURES, AUTISM, ATAXIA, INTERMITTENT HYPERVENTILATION, AND STEREOTYPIC HAND MOVEMENTS. AFTER INITIAL REGRESSION, THE CONDITION STABILIZES AND PATIENTS USUALLY SURVIVE INTO ADULTHOOD.
CC -!- SIMILARITY: CONTAINS (R)GRE(K) MOTIFS AND SPKK MOTIFS THAT HAVE BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
CC -!- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).

```


Q92206:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Methyl-CpG-binding protein 2 (MeCP-2 protein) (MeCP2).
 GN MECP2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=98449942; PubMed=9774669;
 RA Hendrich B., Bird A.;
 RT "Identification and characterization of a family of mammalian methyl-
 CpG binding proteins.";
 RL Mol. Cell. Biol. 18:6538-6547(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99299240; PubMed=10369871;
 RA Coy J.F., Sedlacek Z., Baechner D., Dellus H., Poustka A.;
 RT "A complex pattern of evolutionary conservation and alternative
 polyadenylation within the long 3'-untranslated region of the methyl-
 CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene
 expression.";
 RL Hum. Mol. Genet. 8:1253-1262(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Reichwald K., Thiessen J., Wiehe T., Kioschis P., Straetling W.H.,
 Rostenthal A., Platzer M.;
 RT "Comparative analysis of the methyl CpG binding protein 2 locus in man
 and mouse reveals new untranslated sequences.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBI databases.
 CC -1- FUNCTION: CHROMOSOMAL PROTEIN THAT BINDS TO METHYLATED DNA. IT CAN
 BIND SPECIFICALLY TO A SINGLE METHYL-CPG PAIR. IT IS NOT
 INFLUENCED BY SEQUENCES FLANKING THE METHYL-CPGs. MEDIATES
 TRANSCRIPTIONAL REPRESSION THROUGH INTERACTION WITH HISTONE
 DEACETYLASE AND THE COREPRESSOR SIN3A (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. COLOCALIZED WITH METHYL-CPG IN THE
 GENOME.
 CC -1- SIMILARITY: CONTAINS (R)GRP(K) MOTIFS AND SPK MOTIFS THAT HAVE
 BEEN FOUND TO BIND TO MINOR GROOVE OF AT-RICH DNA.
 CC -1- SIMILARITY: CONTAINS 1 METHYL-BINDING DOMAIN (MBD).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AF072251; AAC68880.1; -
 DR EMBL; AJ132522; CAB46495.1; -
 DR EMBL; AF121351; AAF22116.1; -
 DR EMBL; AF158181; AAF33024.1; -
 DR MGD; MGI:99918; Mecp2.
 DR InterPro: IPR001739; MBD.
 DR InterPro: IPR000637; At_hook.
 DR Pfam; PF01429; MBD; 1.
 DR PRINTS; SM00929; ATHOOK.
 DR SMART; PR00391; MBD; 1.
 DR Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 KW DOMAIN 96 149 MBD.
 FT DOMAIN 277 283 POLY-ALA.
 FT DOMAIN 366 372 POLY-HIS.
 FT DOMAIN 384 391 POLY-PRO.
 FT DOMAIN 440 443 POLY-THR.
 SQ SEQUENCE 484 AA; 52307 MW; 62FD228F0118A49F CRC64;

alignment_scores:

Quality: 58.00 Length: 11
 Ratio: 5.800 Gaps: 0
 Percent Similarity: 90.909 Percent Identity: 72.727

alignment_block:

US-09-674-779-4/rev x MEC2_MOUSE ..

Align seg 1/1 to: MEC2_MOUSE from: 1 to: 484

51 CCAAAATAAGCTACACCATCACCATCACCAT 19

|||||:|||||:|||||:|||||:|||||:|||||

362 ProllyLysGluHisHisHisHisHisHis 372

seq_name: SwissProt_40:SNF1_CANGA

seq_documentation_block:

ID SNF1_CANGA STANDARD; PRT; 611 AA.

AC Q00372;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Carbon catabolite derepressing protein kinase (EC 2.7.1.-).

GN SNF1

OS Candida glabrata (Yeast) (Torulopsis glabrata).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5478;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCCLS84;

RX MEDLINE=97101049; PubMed=8945576;

RA Pether R., Kwon-Chung K.J.;

RT "Disruption of the SNF1 gene abolishes trehalose utilization in the

pathogenic yeast Candida glabrata.";

RL Infect. Immun. 64:5269-5273(1996).

CC -1- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT

INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY

PROTEIN SNF4. COULD PHOSPHORYLATES CAT8 (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY

SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC SNF1 SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; L78130; AAB48642.1; -

DR HSSP; P00518; 2PHK.

DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR002290; Ser_thr_pkinase.

DR Pfam; PF00069; pkinase; 1.

DR SMART; SM00220; S_TKC; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

KW Transferase; Serine/threonine-protein kinase; ATP-binding;

Phosphorylation; Carbohydrate metabolism; Nuclear protein.

FT DOMAIN 6 17 POLY-HIS.

FT DOMAIN 39 290 PROTEIN KINASE.

FT NP_BIND 45 53 ATP (BY SIMILARITY).

FT BINDING 68 68 ATP (BY SIMILARITY).

FT ACT_SITE 161 161 BY SIMILARITY.

FT MOD_RES 194 194 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

SQ SEQUENCE 611 AA; 70049 MW; 89E17812A490CD0 CRC64;

alignment_scores:

Quality: 57.00

Length: 10

Ratio: 6.333 Gaps: 0
Percent Similarity: 90.000 Percent Identity: 80.000

alignment_block:

US-09-674-779-4/rev x SNFL_CANGA ..

Align seg 1/1 to: SNFL_CANGA from: 1 to: 611

48 AAAAATAAGTACACCATCACCACCATCAT 19

2 GluAsnLysGluHisHisHisHisHisHis 11

seq_name: SwissProt_40:DYRA_HUMAN

seq_documentation_block:

ID DYRA_HUMAN STANDARD; PRT: 763 RA.
AC Q13627; Q92810; Q92582; Q9UNW5;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A
DE (EC 2.7.1.-) (Protein kinase minibrain homolog) (MNBH) (HP86) (Dual
DE specificity YAK1-related kinase).
GN DYRK1A OR DYRK OR MNBH OR MNB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=97131512; PubMed=8975710;
RA Song W.J., Sternberg L.R., Kasten-Sportes C., van Keuren M.L.,
RA Chung S.H., Slack A.C., Miller D.E., Glover T.W., Chiang P.W., Lou L.,
RA Kurnit D.W.;
RT "Isolation of human and murine homologues of the Drosophila minibrain
RT gene: human homologue maps to 21q22.2 in the Down syndrome 'critical
RT region'";
RL Genomics 38:331-339(1996).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=97026291; PubMed=8872470;
RA Guimera A.M., Casas C., Pucharcos C., Solans A., Domenech A.,
RA Planas A.M., Ashley J., Lovett M., Estivill X., Pritchard M.A.;
RT "A human homologue of Drosophila minibrain (MNB) is expressed in the
RT neuronal regions affected in Down syndrome and maps to the critical
RT region.";
RL Hum. Mol. Genet. 5:1305-1310(1996).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Fetal brain;
RC MEDLINE=96332410; PubMed=8769099;
RA Shindoh N., Kudoh J., Maeda H., Yamaki A., Minoshima S., Shimizu Y.,
RA Shimizu N.;
RT "Cloning of a human homolog of the Drosophila minibrain/rat Dyrc gene
RT from 'the Down syndrome critical region' of chromosome 21.";
RL Biochem. Biophys. Res. Commun. 225:92-99(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Miki O., Masahira H., Yasufumi M., Naohiko S., Takahiro N.,
RA Hitoshi I., Nobuo N., Toshihiko E., Yoshiyuki S., Misao O.,
RA Ohira M., Seki N., Nagase T., Suzuki E., Nomura N., Ohara O.,
RA Hattori M., Sakaki Y., Eki T., Murakami Y., Saito T., Ichikawa H.,
RA Ohki M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A., VARIANTS F-415; H-681, AND ALTERNATIVE SPLICING.
RX MEDLINE=99263503; PubMed=10329007;
RA Guimera J., Casas C., Estivill X., Pritchard M.;
RT "Human minibrain homologue (MNBH/DYRK1): characterization, alternative
RT splicing, differential tissue expression, and overexpression in Down
RL syndrome.";
RL Genomics 57:407-418(1999).
CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING

CC NUCLEAR FUNCTIONS OF CELL PROLIFERATION. PHOSPHORYLATES SERINES,
CC THREONINES AND TYROSINES RESIDUS IN ITS SEQUENCE AND IN EXOGENOUS
CC SUBSTRATES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; A LONG FORM (SHOWN
CC HERE), 1, 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGHEST LEVELS IN SKELETAL MUSCLE,
CC TESTIS, FETAL LUNG AND FETAL KIDNEY.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING CENTRAL NERVOUS
CC SYSTEM.
CC -!- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
CC -!- DISEASE: OVEREXPRESSED 1.5-FOLD IN FETAL DOWN SYNDROME BRAIN.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MNB/DYRK SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; U58496; AAC50939.1; -
CC EMBL; U52373; AAB18639.1; -
CC EMBL; D85759; BAAL2866.1; -
CC EMBL; D86550; BAAL3110.1; -
CC EMBL; AF108830; AAD31169.1; -
CC HSP; P27703; IERK.
CC MIM; 600855; -
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 2.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
CC ATP-binding; Nuclear protein; Phosphorylation; Alternative splicing;
CC Polymorphism.
KW DOMAIN 117 134 BIPARTITE NUCLEAR LOCALIZATION SIGNAL
FT (POTENTIAL).
FT PROTEIN KINASE.
FT NP_BIND 159 479
FT BINDING 165 173 ATP (BY SIMILARITY).
FT ACT_SITE 188 188 ATP (BY SIMILARITY).
FT DOMAIN 287 287 BY SIMILARITY.
FT DOMAIN 509 515 POLY-SER.
FT DOMAIN 599 602 POLY-HIS.
FT DOMAIN 607 619 POLY-HIS.
FT DOMAIN 656 672 SER/THR-RICH.
FT DOMAIN 664 671 POLY-SER.
FT MOD_RES 219 219 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 319 319 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 321 321 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT VARSPPLIC 70 78 MISSING (IN ISOFORM 1).
FT VARSPPLIC 516 529 GSGSTSGSGRS -> GASATSCSWLYRH (IN
FT ISOFORM 2).
FT VARSPPLIC 530 763 MISSING (IN ISOFORM 2).
FT VARSPPLIC 516 540 GSGSTSGSGRSDPHQHSGG -> VEQHWMPGAFRM
FT TVSFLEVDVVPV (IN ISOFORM 3).
FT VARSPPLIC 541 763 MISSING (IN ISOFORM 3).
FT VARSPPLIC 559 584 RQPPAPLWGSQTAPQTVVETHPV -> SSHVVHLLVSP
FT VARSPPLIC 585 763 AILWWSGTGQVPLE (IN ISOFORM 4).
FT VARIANT 415 415 MISSING (IN ISOFORM 4).
FT Y -> F.
FT Q -> H.
FT Q -> H.
FT G -> A (IN REF. 1).
FT N -> S (IN REF. 1).
FT S -> P (IN REF. 1).
FT O -> R (IN REF. 1).
FT K -> N (IN REF. 1).
FT CONFLICT 397 397

FT CONFLICT 592 592 A -> G (IN REF. 1).
 FT CONFLICT 679 679 A -> P (IN REF. 1).
 SQ SEQUENCE 763 AA; 85584 MW; 7C3A52A3CB04FB5 CRC64;

alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x DYRA_HUMAN ..

Align seg 1/1 to: DYRA_HUMAN from: 1 to: 763

60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||||
 590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisHisGlyAsnSerSe 606

37 ACACCATCACCACCAT 19
 ||||| ||||| ||||| |||||

606 rHisHisHisHisHisHis 612

seq_name: SwissProt_40:DYRA_MOUSE

seq_documentation_block:

ID DYRA_MOUSE STANDARD; PRT; 763 AA.

AC Q61214;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A
 DE (EC 2.7.1.1-) (Protein kinase minibrain homolog) (MNBH) (MP86) (Dual
 DE specificity YAKI-related kinase).
 GN DYRK1A OR DYRK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER / NIH;
 RX MEDLINE=97131512; PubMed=8975710;
 RA Song W.J., Sternberg L.R., Kasten-Sportes C., van Keuren M.L.,
 RA Chung S.H., Slack A.C., Miller D.E., Glover T.W., Chiang P.W., Lou L.,
 RA Kurnit D.W.;
 RA "Isolation of human and murine homologues of the Drosophila minibrain
 RT gene: human homologue maps to 21q22.2 in the Down syndrome 'critical
 RT region'";
 RL Genomics 38:331-339(1996).

[2]
 RP SUBCELLULAR LOCATION, AND SUBUNIT.
 RX MEDLINE=97224401; PubMed=9070862;
 RA Song W.J., Chung S.H., Kurnit D.M.;
 RT "The murine Dyk protein maps to chromosome 16, localizes to the
 RT nucleus, and can form multimers";
 RL Biochem. Biophys. Res. Commun. 231:640-644(1997).
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING
 CC NUCLEAR FUNCTIONS OF CELL PROLIFERATION. PHOSPHORYLATES SERINES,
 CC THREONINES AND TYROSINES RESIDUES IN ITS SEQUENCE AND IN EXOGENOUS
 CC SUBSTRATES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MNB/DYRK SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U58497; AAC52994.1; ..
 DR HSP; P27703; IERK
 DR MGD; MGI:1330299; Dyrla.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 2.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_SF; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Nuclear protein; Phosphorylation.
 FT BIPARTITE NUCLEAR LOCALIZATION SIGNAL
 FT DOMAIN 117 134
 FT FT
 FT DOMAIN 159 479
 FT NP_BIND 165 173
 FT BINDING 188 188
 FT ACT_SITE 287 287
 FT DOMAIN 509 515
 FT DOMAIN 599 602
 FT DOMAIN 607 619
 FT DOMAIN 656 672
 FT DOMAIN 664 671
 FT MOD_RES 219 219
 FT MOD_RES 319 319
 FT MOD_RES 321 321
 SQ SEQUENCE 763 AA; 85494 MW; E117DDD6C5E8C74F CRC64;

alignment_scores:

Quality: 56.50 Length: 23
 Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x DYRA_MOUSE ..

Align seg 1/1 to: DYRA_MOUSE from: 1 to: 763

60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||||
 590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisHisGlyAsnSerSe 606

37 ACACCATCACCACCAT 19
 ||||| ||||| ||||| |||||

606 rHisHisHisHisHisHis 612

seq_name: SwissProt_40:DYRA_MOUSE

seq_documentation_block:

ID DYRA_MOUSE STANDARD; PRT; 763 AA.

AC Q63470;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dual-specificity tyrosine-phosphorylation regulated kinase 1A
 DE (EC 2.7.1.1-) (Protein kinase minibrain homolog) (MNBH) (RP86) (Dual
 DE specificity YAKI-related kinase).
 GN DYRK1A OR DYRK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A., AND MUTAGENESIS OF TYR-219.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
 RX MEDLINE=96216443; PubMed=8631952;
 RA Kenrup H., Becker W., Heukelbach J., Wilmes A., Schuermann A.,
 RA Huppertz C., Raulainen H., Joost H.-G.;
 RA "Dyrk, a dual specificity protein kinase with unique structural
 RT features whose activity is dependent on tyrosine residues between
 RT subdomains VII and VIII.";

RL J. Biol. Chem. 271:3488-3495(1996).
 RN [2]
 RP REVISIONS.
 RA Kentrup H.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98421512; PubMed=9748265;
 RA Becker W., Weber Y., Wetzel K., Eirmbter K., Tejedor F.J.,
 RA Joost H.-G.;
 RT "Sequence characteristics, subcellular localization, and substrate
 RT specificity of DYRK-related kinases, a novel family of dual
 RT specificity protein kinases.";
 RL J. Biol. Chem. 273:25893-25902(1998).
 CC -!- FUNCTION: MAY PLAY A ROLE IN A SIGNALING PATHWAY REGULATING
 CC NUCLEAR FUNCTIONS OF CELL PROLIFERATION. PHOSPHORYLATES SERINES,
 CC THREONINES AND TYROSINES RESIDUES IN ITS SEQUENCE AND IN EXOGENOUS
 CC SUBSTRATES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -!- PFM: AUTOPHOSPHORYLATED ON TYR RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MNE/DYRK SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X79769; CAA56164.1; --
 DR HSPSP; P27703; 1ERK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART: SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_SP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Nuclear protein; BIPARTITE NUCLEAR LOCALIZATION SIGNAL
 FT DOMAIN 117 134
 FT 159 479
 FT NP_BIND 165 173
 FT BINDING 188 188
 FT ACT_SITE 287 287
 FT DOMAIN 509 515
 FT 599 602
 FT DOMAIN 607 619
 FT 656 672
 FT DOMAIN 664 671
 FT 219 219
 FT MOD_RES 319 319
 FT MOD_RES 321 321
 FT MOD_RES 70 78
 FT VARSPLIC 219
 FT MUTAGEN 219
 FT 763 AA; 85541 MW; CB5EC7EC4C1F9A47 CRC64;
 SQ SEQUENCE

alignment_scores:
 Quality: 56.50 Length: 23

Ratio: 4.708 Gaps: 1
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x DYRA_RAT

Align seg 1/1 to: DYRA_RAT from: 1 to: 763
 60 CATGCTCTTACCAAAA.....AATAAGCT 38
 ||||| |||:::
 590 HisValAlaProGlnGlnAsnAlaLeuHisHisHisGlyAsnSer 606
 ||||| |||||
 37 ACACATCATCACCATCACCAT 19
 ||||| |||||
 606 rHisHisHisHisHisHis 612
 seq_name: SwissProt_40:PC_DROME
 seq_documentation_block:
 ID PC_DROME STANDARD; PRT; 390 AA.
 AC P26017; Q9VP49;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Polycarb protein.
 DE PC OR C67618.
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=91095442; PubMed=1898775;
 RA Paro R., Hogness D.S.;
 RT "The Polycarb protein shares a homologous domain with a
 RT heterochromatin-associated protein of Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:263-267(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Fodor C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:  
The genome sequence of Drosophila melanogaster." ;  
RL Science 287:2185-2195(2000).  
CC -!- FUNCTION: REGENES. PC MAY BE INVOLVED IN THE SEGMENT-SPECIFIC REPRESSON OF HOMEOTIC  
CC SELECTOR GENES. PC MAY BE INVOLVED IN THE STABLE TRANSMISSION OF A  
CC DETERMINED STATE BY ITS EFFECTS ON CHROMATIN STRUCTURE. PROMOTES  
CC LOCUS-SPECIFIC CHROMATIN COMPACTION.  
CC -!- DEVELOPMENTAL STAGE: REQUIRED DURING THE ENTIRE LARVAL PERIOD FOR  
CC NORMAL ADULT DEVELOPMENT. IT IS FOUND IN ALMOST ALL CELLS AND  
CC TISSUES THROUGHOUT GASTRULATION AND ORGANOGENESIS THOUGH AT A MUCH  
CC LOWER LEVEL THAN IN EARLY SYNCTIAL STAGES.  
CC -!- SIMILARITY: CONTAINS 1 'CHROMO' DOMAIN.  
  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
EMBL; X55702; CAA39229.1; -.  
DR EMBL; AE003594; AAF51707.1; -.  
PIR; A38565; A38565.  
DR HSSP; P23197; LAP0.  
TRANSPAC; T00693; -.  
FlyBase; FBgn0003042; Pc.  
InterPro; IPR000953; Chromo.  
PFam; PF00385; chromo; 1.  
PRINTS; PR00504; CHROMODOMAIN.  
SMART; SM00298; CHROMO; 1.  
PROSITE; PS00598; CHROMO_1; 1.  
PROSITE; PS00013; CHROMO_2; 1.  
Nuclear protein..  
FT DOMAIN 26 84 CHROMO.  
FT DOMAIN 134 143 POLY-HIS.  
FT DOMAIN 160 167 POLY-HIS.  
FT SIMILAR 351 380 TO MODIFIER 3 PROTEIN (AA 488-517).  
SEQUENCE 390 AA; 43976 MW; 5DB24AE4B326C3B9 CRC64;  
  
alignment_scores:  
Quality: 56.00 Length: 10  
Ratio: 6.222 Gaps: 0  
Percent Similarity: 90.000 Percent Identity: 80.000  
  
alignment_block:  
US-09-674-779-4/rev x PC_DROME ..  
Align seg 1/1 to: PC_DROME from: 1 to: 390  
  
seq_name: SwissProt_40:FXGA_HUMAN  
seq_documentation_block:  
ID FXGA_HUMAN STANDARD; PRT; 469 AA.  
AC P55316;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Forkhead box protein GlA (Forkhead-related protein FKHL2)  
DE (Transcription factor BF-2) (Brain factor 2) (BF2) (HFK2).  
GN FOXGIA OR FKHL2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;
```


FT DOMAIN 124 127 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN_FING 381 420 RING-TYPE.
FT DOMAIN 357 476 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 477 688 PRO-RICH.
FT DOMAIN 689 834 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 856 895 UBA.
FT MOD_RES 700 700 PHOSPHORYLATION.
FT MOD_RES 774 774 PHOSPHORYLATION.
SQ SEQUENCE 906 AA; 99646 MW; 7D686B050204AD8F CRC64;

alignment_scores:
Quality: 56.00 Length: 12
Ratio: 5.600 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 66.667

alignment_block:

US-09-674-779-4/rev x CBL_HUMAN ..

Align seg 1/1 to: CBL_HUMAN from: 1 to: 906

36 CACCATCACCATCACCATTAAAGTCGACCCCTCGCT 1
|||||
36 HishishishishishishisLeuSerProHisPro 47

seq_name: SwissProt_40:URBL_USTMA

seq_documentation_block:

ID URBL_USTMA STANDARD; PRT; 950 AA.

AC P40349;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Siderophore biosynthesis regulatory protein URBS1.
GN URBS1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=518;
RX MEDLINE=94019380; PubMed=8413298;
RA Volzard C.P., Wang J., McEvoy J.L., Xu P., Leong S.A.;
RT "urbs1, a gene regulating siderophore biosynthesis in Ustilago
maydis, encodes a protein similar to the erythroid transcription
factor GATA-1.";
RL Mol. Cell. Biol. 13:7091-7100(1993).
CC -!- FUNCTION: INVOLVED IN THE REGULATION OF SECRETED PERRICHROME-TYPE
SIDEROPHORES. ACTS DIRECTLY OR INDIRECTLY TO REPRESS THE
BIOSYNTHESIS OF SIDEROPHORES.
CC -!- SUBCELLULAR LOCATION: Nucleus (Probable).
CC -!- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80547; AAB05617.1; -.
CC PIR; S27473; S27473.
CC HSSP; P17679; 1GNF.
CC TRANSFAC; T02406; -.
CC InterPro; IPR000679; ZnF_GATA.
CC InterPro; IPR001164; ZnF_GCS.
CC Pfam; PF00320; GATA; 2.
CC PRINTS; PR00619; GATAGNFINGER.
CC SMART; SM00401; ZnF_GATA; 2.
CC PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
CC PROSITE; PS50114; GATA_ZN_FINGER_2; 2.

KW DNA-binding; Zinc-finger; Transcription regulation; Repressor;
KW Nuclear protein.
FT ZN_FING 338 362 GATA-TYPE 1.
FT ZN_FING 482 506 GATA-TYPE 2.
FT DOMAIN 24 27 POLY-ALA.
FT DOMAIN 28 36 POLY-SER.
FT DOMAIN 487 490 POLY-THR.
FT DOMAIN 753 780 HIS-RICH.
FT DOMAIN 902 906 POLY-SER.
SQ SEQUENCE 950 AA; 101427 MW; F969C6DA09A78C12 CRC64;

alignment_scores:
Quality: 56.00 Length: 13
Ratio: 5.600 Gaps: 0
Percent Similarity: 76.923 Percent Identity: 61.538

alignment_block:

US-09-674-779-4/rev x URBL_USTMA ..

Align seg 1/1 to: URBL_USTMA from: 1 to: 950

57 GCTTACCAAAATAAGCTACCATCACCATCACCAT 19
|||||
750 ValValProHishishishishishishishisHis 762

seq_name: SwissProt_40:CCAB_DISOM

seq_documentation_block:

ID CCAB_DISOM STANDARD; PRT; 2326 AA.

AC P56698;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable voltage-dependent N-type calcium channel alpha-1B subunit
(DOE-4).
OS Discopyle ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosqualea; Pristiogaster; Batoidae;
OC Torpediniformes; Narcinoidei; Narcinidae; Discopyge.
OX NCBI_TaxID=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Electric lobe;
RX MEDLINE=93248175; PubMed=7683405;
RA Horne W.A., Ellinor P.T., Inman I., Zhou M., Tsien R.W., Schwarz T.L.;
RT "Molecular diversity of Ca2+ channel alpha 1 subunits from the marine
ray Discopyle ommata.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3787-3791(1993).
CC -!- FUNCTION: THE ISOFORM ALPHA-1B GIVES RISE TO N-TYPE CALCIUM
CURRENTS. N-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE
ACTIVATED" (HVA) GROUP (BY SIMILARITY).
CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY
SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY
ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN THE ELECTRIC LOBE THAN
IN THE FOREBRAIN.
CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -!- PTM: PHOSPHORYLATED IN VITRO BY CAM-KINASE II, CAPK, PKC AND GPK
(BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS

CC CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L12532; NOT_ANNOTATED_CDS.

DR InterPro: IPR002077; Ca_channel.

DR InterPro: IPR002111; Cat_channel_trpl.

DR InterPro: IPR000636; Cation_chan_non_lig.

DR InterPro: IPR001682; Channel_pore_Ca_Na.

DR Pfam: PF00520; Ion_trans_4.

DR PRINTS: PR00167; CACHANNEL.

KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;

KW Calcium channel; Glycoprotein; Repeat; Multigene family;

KW Calcium-binding; Phosphorylation; Alternative splicing.

FT REPEAT 75 351 I.

FT REPEAT 458 702 II.

FT REPEAT 1134 1416 III.

FT REPEAT 1453 1708 IV.

FT DOMAIN 89 188 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 108 107 S1 OF REPEAT I (POTENTIAL).

FT DOMAIN 108 125 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 126 145 S2 OF REPEAT I (POTENTIAL).

FT DOMAIN 146 156 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 157 176 S3 OF REPEAT I (POTENTIAL).

FT DOMAIN 177 180 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 181 199 S4 OF REPEAT I (POTENTIAL).

FT DOMAIN 200 218 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 219 238 S5 OF REPEAT I (POTENTIAL).

FT DOMAIN 239 323 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 324 348 S6 OF REPEAT I (POTENTIAL).

FT DOMAIN 349 472 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 473 491 S1 OF REPEAT II (POTENTIAL).

FT DOMAIN 492 506 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 507 526 S2 OF REPEAT II (POTENTIAL).

FT DOMAIN 527 534 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 535 552 S3 OF REPEAT II (POTENTIAL).

FT DOMAIN 553 563 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 564 582 S4 OF REPEAT II (POTENTIAL).

FT DOMAIN 583 601 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 602 621 S5 OF REPEAT II (POTENTIAL).

FT DOMAIN 622 674 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 675 699 S6 OF REPEAT II (POTENTIAL).

FT DOMAIN 700 1148 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1149 1166 S1 OF REPEAT III (POTENTIAL).

FT DOMAIN 1167 1182 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1183 1202 S2 OF REPEAT III (POTENTIAL).

FT DOMAIN 1203 1214 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1215 1233 S3 OF REPEAT III (POTENTIAL).

FT DOMAIN 1234 1243 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1244 1262 S4 OF REPEAT III (POTENTIAL).

FT DOMAIN 1263 1281 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1282 1301 S5 OF REPEAT III (POTENTIAL).

FT DOMAIN 1302 1388 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1389 1413 S6 OF REPEAT III (POTENTIAL).

FT DOMAIN 1414 1468 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1469 1487 S1 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1488 1502 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1503 1522 S2 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1523 1530 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1531 1549 S3 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1550 1558 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1559 1577 S4 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1578 1596 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1597 1616 S5 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1617 1680 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1681 1705 S6 OF REPEAT IV (POTENTIAL).

FT DOMAIN 1706 2326 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 1869 1873 POLY-GLN.

FT DOMAIN 2040 2046 POLY-HIS.

FT DOMAIN 371 398 BINDING TO THE BETA SUBUNIT (BY SIMILARITY).

FT SITE 306 306 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 653 653 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 1362 1362 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT SITE 1650 1650 CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).

FT MOD_RES 1716 1716 PHOSPHORYLATION (BY CAPK) (POTENTIAL).

FT CA_BIND 1734 1745 BY SIMILARITY.

FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1558 1558 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 406 406 D -> DQGLGIIEPEQKPEDIQSVY (IN ISOPFORM 2).

SQ SEQUENCE 2326 AA; 264515 MW; D5DEAA09E819B6B CRC64;

alignment_scores:

Quality: 55.50 Length: 23

Ratio: 4.625 Gaps: 1

Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:

US-09-674-779-4/rev x CCAB_DISOM ..

Align seg 1/1 to: CCAB_DISOM from: 1 to: 2326

60 CATGCTCTTACCAAAAATAAGCTA..... 37

|||||:|||||:|||||

2023 HisValMetProAspTyrSerLeuGluArgValleProValGlnMetPr 2039

36 .CACCATCACCATCACCAT 19

|||||:|||||:|||||

2039 oHisHisHisHisHisHis 2045

seq_name: SwissProt_40:MAFL_MOUSE

seq_documentation_block:

ID MAFL_MOUSE STANDARD; PRT; 323 AA.

AC P54841;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Transcription factor MAFL (Segmentation protein KR) (Kreisler).

GN MAFB OR MAFL OR KRML.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95094266; PubMed=8001130;

RA Cordes S.P., Barsh G.S.;

RT "The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor.";

RL Cell 79:1025-1034(1994).

CC -!- FUNCTION: MAY PLAY AN EARLY ROLE IN AXIAL PATTERNING (HINDBRAIN SEGMENTATION).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.

CC -!- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND

CC IN THE CAUDAL HINDBRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),

CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT

CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY

CC THROUGH R6.

CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -


```

ID Q95LG8 PRELIMINARY; PRT; 486 AA.
AC Q95LG8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE METHYL-CPG-BINDING PROTEIN 2.
GN MECP2.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Muramatsu S.;
RT "Excessive hand-wringing in a mmp2-treated monkey.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF295597; AAK97131.1; -.
SQ SEQUENCE 486 AA; 52426 MW; 3471B61D90D92A7D CRC64;

alignment_scores:
  Quality: 59.00 Length: 15
  Ratio: 4.917 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 60.000

alignment_block:
US-09-674-779-4/rev x Q95LG8 ..
Align seg 1/1 to: Q95LG8 from: 1 to: 486

51 CCAAAATAAGTACACCATCACCATCACCATTAGTCGACCCCT 7
|||||:|||||:|||||:|||||:|||||:|||||:
362 ProLysLysGluHisHisHisHisSerGluSerPro 376

seq_name: sp_invertebrate:Q9VZ60

seq_documentation_block:
ID Q9VZ60 PRELIMINARY; PRT; 1128 AA.
AC Q9VZ60;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG2186 PROTEIN.
GN CG2186.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Arl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003484; AAF47967.1; -.
DR FlyBase; FBgn0030243; CG2186.
SQ SEQUENCE 1128 AA; 122705 MW; EA6FA5F44B8D645C CRC64;

alignment_scores:
  Quality: 59.00 Length: 13
  Ratio: 5.364 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 61.538

alignment_block:
US-09-674-779-4/rev x Q9VZ60 ..
Align seg 1/1 to: Q9VZ60 from: 1 to: 1128

60 CATGCTTTACAAAATAAGCTACACCATCACCATCAC 22
|||||:|||||:|||||:|||||:|||||:
910 HisLeuAlaProSerSerLysGlnHisHisHisHis 922

seq_name: sp_invertebrate:Q95Z58

seq_documentation_block:
ID Q95Z58 PRELIMINARY; PRT; 1266 AA.
AC Q95Z58;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KRUEPPEL-LIKE PROTEIN.
GN KRUEPPEL.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=3D7;
RC Boschet C.J., Vaquero C.M.;
RT "Plasmodium falciparum krueppel-like gene.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ318337; CAC42230.1; -.
SQ SEQUENCE 1266 AA; 151646 MW; BBD631DE99DEFCC8 CRC64;

alignment_scores:
  Quality: 59.00 Length: 16
  Ratio: 4.917 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
US-09-674-779-4/rev x Q95Z58 ..
Align seg 1/1 to: Q95Z58 from: 1 to: 1266

48 AAAATAAGCTACACCATCACCATCACCATTAGTCGACCCCT 1
|||||:|||||:|||||:|||||:|||||:

```

506 LysAsnAsnLysHisHisHisHisHisHisLeuLysTyrIlePro 521

```
seq_name: sp_human:Q96F26
seq_documentation_block:
ID Q96F26 PRELIMINARY; PRT; 454 AA.
AC Q96F26;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3534143) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
FL Submitted (JUL-2001) to the EMBL/GenBank/DBDJ databases.
DR EMBL; BC011695; AAL11695.1; -.
FT NON_TER
SQ SEQUENCE 454 AA; 50269 MW; 11FC856EF3EEF24E CRC64;

alignment_scores:
Quality: 58.00 Length: 14
Ratio: 5.273 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:
US-09-674-779-4/rev x Q96F26 ..
Align seg 1/1 to: Q96F26 from: 1 to: 454
60 CATGCTTACCAAAATAAGCTACACATCACCATCACCAT 19
||||| :|:|:| | | | | | | | | | | | | | |
358 HisValAspSerThrMetLeuHisHisHisHisHis 371

seq_name: sp_vertebrate:Q9YGC6
seq_documentation_block:
ID Q9YGC6 PRELIMINARY; PRT; 467 AA.
AC Q9YGC6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE METHYL-CPG-BINDING PROTEIN MECP2.
GN MECP2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98282101; PubMed=9620779;
RA Jones P.L., Veenstra G.J.C., Wade P.A., Vermaak D., Kass S.U.,
RA Landsberger N., Strouboulis J., Wolfe A.P.;
RT "Methylated DNA and Mecp2 recruit histone deacetylase to repress
transcription";
RL Nat. Genet. 19:187-191(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kass S.U., Strouboulis J., Wolfe A.P.;
RT "Xenopus Mecp2";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AF106951; AAD03736.1; -.
DR EMBL; AF051768; AAD02651.1; -.
DR InterPro; IPR001739; MBD.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00391; MBD; 1.
SQ SEQUENCE 467 AA; 51757 MW; 5D3A719A595E560BC CRC64;
```

```
alignment_scores:
Quality: 58.00 Length: 14
Ratio: 4.833 Gaps: 1
Percent Similarity: 85.714 Percent Identity: 71.429

alignment_block:
US-09-674-779-4/rev x Q9YGC6 ..
Align seg 1/1 to: Q9YGC6 from: 1 to: 467
54 TTACCAAAAAATAAGCTA.....CACATCACCATCACCAT 19
||||| :|:|:| | | | | | | | | | | | | | |
356 LeuProLysLysGluLeuGlnGlnHisHisHisHisHis 369

seq_name: sp_human:Q9HCK8
seq_documentation_block:
ID Q9HCK8 PRELIMINARY; PRT; 1417 AA.
AC Q9HCK8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1564 PROTEIN (FRAGMENT).
GN KIAA1564.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046784; BAB13390.1; -.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; Helicase; 1.
KW ATP-binding; Helicase.
FT NON_TER
SQ SEQUENCE 1417 AA; 161268 MW; 199854F63D8BC6AE CRC64;

alignment_scores:
Quality: 58.00 Length: 14
Ratio: 5.273 Gaps: 0
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:
US-09-674-779-4/rev x Q9HCK8 ..
Align seg 1/1 to: Q9HCK8 from: 1 to: 1417
60 CATGCTTACCAAAATAAGCTACACATCACCATCACCAT 19
||||| :|:|:| | | | | | | | | | | | | | |
1321 HisValAspSerThrMetLeuHisHisHisHisHis 1334

seq_name: sp_vertebrate:Q9PUM3
seq_documentation_block:
ID Q9PUM3 PRELIMINARY; PRT; 2146 AA.
AC Q9PUM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB4 VARIANT.
GN CACHALB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```



```

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION NEURONS;
RX MEDLINE=20044702; PubMed=10574919;
RA Lu Q., Dunlap K.;
RT "Cloning and functional expression of novel N-type Ca(2+) channel
RT variants.";
RL J. Biol. Chem. 274:34566-34575(1999).
DR EMBL; AF173015; AAD5182.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR002077; Ca_channel.
DR InterPro; IPR001682; Channel_pore_Ca_Na.
DR InterPro; IPR001696; Na_channel.
DR Pfam; PF00520; ion_trans; 4.
DR PRINTS; PR00167; CACHANNEL.
DR PRINTS; PR00170; NACHANNEL.
SQ SEQUENCE 2146 AA; 244650 MW; 281BBECAC494AE49 CRC64;

alignment_scores:
  Quality: 58.00 Length: 12
  Ratio: 5.273 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
US-09-674-779-4/rev x Q9PUM3
Align seg 1/1 to: Q9PUM3 from: 1 to: 2146
..
54 TTACCAAAAATAAGCTACACCATCACCATCACCAT 19
:::||||| :::::::::::|||||
2036 ValProThrAspGlnValHisHisHisHis 2047

seq_name: sp_vertebrate:Q9PW44
seq_documentation_block:
ID Q9PW44 PRELIMINARY; PRT; 2157 AA.
AC Q9PW44;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB8 VARIANT.
GN CACHA1B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION NEURONS;
RX MEDLINE=20044702; PubMed=10574919;
RA Lu Q., Dunlap K.;
RT "Cloning and functional expression of novel N-type Ca2+ Channel
RT Variants.";
RL J. Biol. Chem. 0:0-0(1999).
DR EMBL; AF173017; AAD51820.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR002077; Ca_channel.
DR InterPro; IPR001682; Channel_pore_Ca_Na.
DR InterPro; IPR001696; Na_channel.
DR Pfam; PF00520; ion_trans; 4.
DR PRINTS; PR00167; CACHANNEL.
DR PRINTS; PR00170; NACHANNEL.
SQ SEQUENCE 2171 AA; 247055 MW; 4023DD3D1AFA7EC3 CRC64;

alignment_scores:
  Quality: 58.00 Length: 12
  Ratio: 5.273 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
US-09-674-779-4/rev x Q9PW46
Align seg 1/1 to: Q9PW46 from: 1 to: 2171
..
54 TTACCAAAAATAAGCTACACCATCACCATCACCAT 19
:::||||| :::::::::::|||||
2061 ValProThrAspGlnValHisHisHisHisHis 2072

seq_name: sp_vertebrate:Q9PUM5
seq_documentation_block:
ID Q9PUM5 PRELIMINARY; PRT; 2182 AA.
AC Q9PUM5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB2 VARIANT.
GN CACHA1B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

```

Ratio: 5.273 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
US-09-674-779-4/rev x Q9PW44
Align seg 1/1 to: Q9PW44 from: 1 to: 2157
..
54 TTACCAAAAATAAGCTACACCATCACCATCACCAT 19
:::||||| :::::::::::|||||
2047 ValProThrAspGlnValHisHisHisHisHis 2058

seq_name: sp_vertebrate:Q9PW46
seq_documentation_block:
ID Q9PW46 PRELIMINARY; PRT; 2171 AA.
AC Q9PW46;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB6 VARIANT.
GN CACHA1B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION NEURONS;
RA Lu Q., Dunlap K.;
RT "Cloning and Functional Expression of Novel N-type Ca2+ Channel
RT Variants.";
RL J. Biol. Chem. 0:0-0(1999).
DR EMBL; AF173017; AAD51820.1; -.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR002077; Ca_channel.
DR InterPro; IPR001682; Channel_pore_Ca_Na.
DR InterPro; IPR001696; Na_channel.
DR Pfam; PF00520; ion_trans; 4.
DR PRINTS; PR00167; CACHANNEL.
DR PRINTS; PR00170; NACHANNEL.
SQ SEQUENCE 2171 AA; 247055 MW; 4023DD3D1AFA7EC3 CRC64;

alignment_scores:
  Quality: 58.00 Length: 12
  Ratio: 5.273 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:
US-09-674-779-4/rev x Q9PW46
Align seg 1/1 to: Q9PW46 from: 1 to: 2171
..
54 TTACCAAAAATAAGCTACACCATCACCATCACCAT 19
:::||||| :::::::::::|||||
2061 ValProThrAspGlnValHisHisHisHisHis 2072

seq_name: sp_vertebrate:Q9PUM5
seq_documentation_block:
ID Q9PUM5 PRELIMINARY; PRT; 2182 AA.
AC Q9PUM5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB2 VARIANT.
GN CACHA1B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

```

Wed Jul 31 08:20:12 2002

Percent Similarity: 91.667 Ratio: 5.273 Gaps: 0
Percent Identity: 58.333

alignment_block:

US-09-674-779-4/rev x Q9PUM4 ..

Align seg 1/1 to: Q9PUM4 from: 1 to: 2321

54 TTACCAAAAATAAGCTACACCATCACCATCACCATCACCAT 19
:::||||| ::::::::::::::::::::|||
2036 ValProThrAspGlnValHisHisHisHisHis 2047

OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION NEURONS;
RX MEDLINE=20044702; PubMed=10574919;
RA Lu Q., Dunlap K.;
RT "Cloning and functional expression of novel N-type Ca(2+) channel
variants";
RL J. Biol. Chem. 274:34566-34575(1999).
DR EMBL: AF173013; AAD51816.1; -;
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR002111; Cat_channel_TripL.
DR InterPro: IPR002077; Ca_channel.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR001696; Na_channel.
DR Pfam: PF00520; ion_trans_4.
DR PRINTS; PR00167; CACHANNEL.
DR PRINTS; PR00170; NACHANNEL.
SQ SEQUENCE 2182 AA; 248259 MW; 1C34F9DF5991783E CRC64;

alignment_scores:
Quality: 58.00 Length: 12
Ratio: 5.273 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 58.333

alignment_block:

US-09-674-779-4/rev x Q9PUM5 ..

Align seg 1/1 to: Q9PUM5 from: 1 to: 2182

54 TTACCAAAAATAAGCTACACCATCACCATCACCATCACCAT 19
:::||||| ::::::::::::::::::::|||
2072 ValProThrAspGlnValHisHisHisHisHis 2083

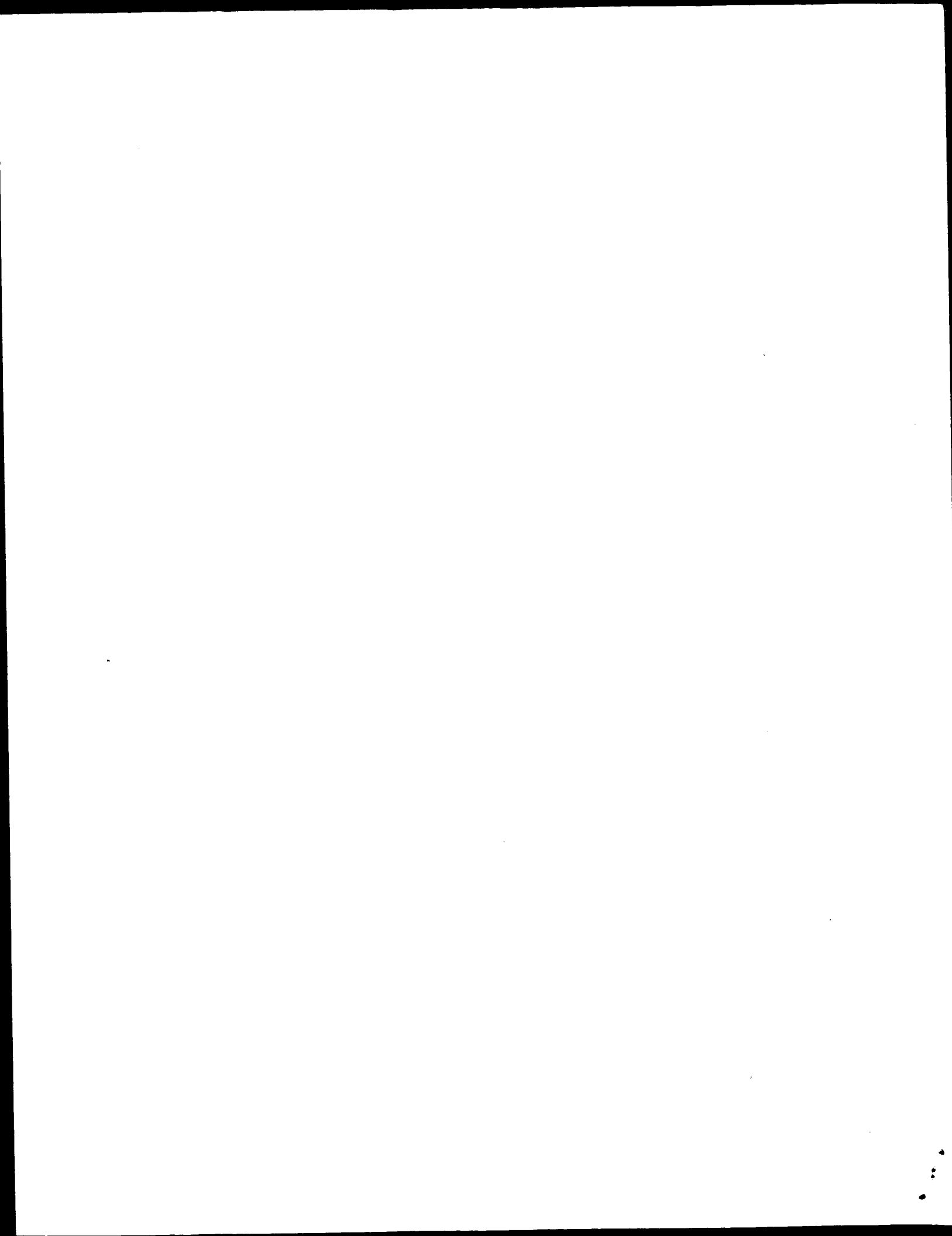
seq_name: sp_vertebrate:Q9PUM4

seq_documentation_block:
ID Q9PUM4 PRELIMINARY; PRT; 2321 AA.
AC Q9PUM4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE N-TYPE CALCIUM CHANNEL ALPHA-1B CDB3 VARIANT.
GN CACHA1B.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=DORSAL ROOT GANGLION NEURONS;
RX MEDLINE=20044702; PubMed=10574919;
RA Lu Q., Dunlap K.;
RT "Cloning and functional expression of novel N-type Ca(2+) channel
variants";
RL J. Biol. Chem. 274:34566-34575(1999).
DR EMBL: AF173014; AAD51817.1; -;
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR002111; Cat_channel_TripL.
DR InterPro: IPR002077; Ca_channel.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR001696; Na_channel.
DR Pfam: PF00520; ion_trans_4.
DR PRINTS; PR00167; CACHANNEL.
DR PRINTS; PR00170; NACHANNEL.
SQ SEQUENCE 2321 AA; 262972 MW; E6FB221909BFD20B CRC64;

alignment_scores:
Quality: 58.00 Length: 12

Wed Jul 31 08:20:12 2002

us-09-674-779-4.rspt



OM of: US-09-674-779-4 to: A_Geneseq_032802:* out_format : pfs

Date: Jul 30, 2002 4:32 PM

About: Results produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framet.n2p.model -DEV=xlp  
-Q=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151835_8501/app_query.fasta_1.116  
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=oligo.rag  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000  
-LOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000  
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000  
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200  
-THR_SCORE=quality -THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09674779_@cgn1_1_88 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-674-779-4  
Query length: 60  
Database: A_Geneseq_032802:*  
Database sequences: 747574  
Database length: 111073796  
Search time (sec): 53.240000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig zScore EScore Len ! Documentation ..



OM of: US-09-674-779-4 to: Issued_Patents_AA:* out_format : pfs

Date: Jul 30, 2002 4:32 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet.n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09674779/runat_30072002_151836_8512/app_query.fasta_1.116
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=oligo.ra1
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200
-THR_SCORE=quality -THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_29 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-674-779-4

Query length: 60

Database: Issued_Patents_AA:*

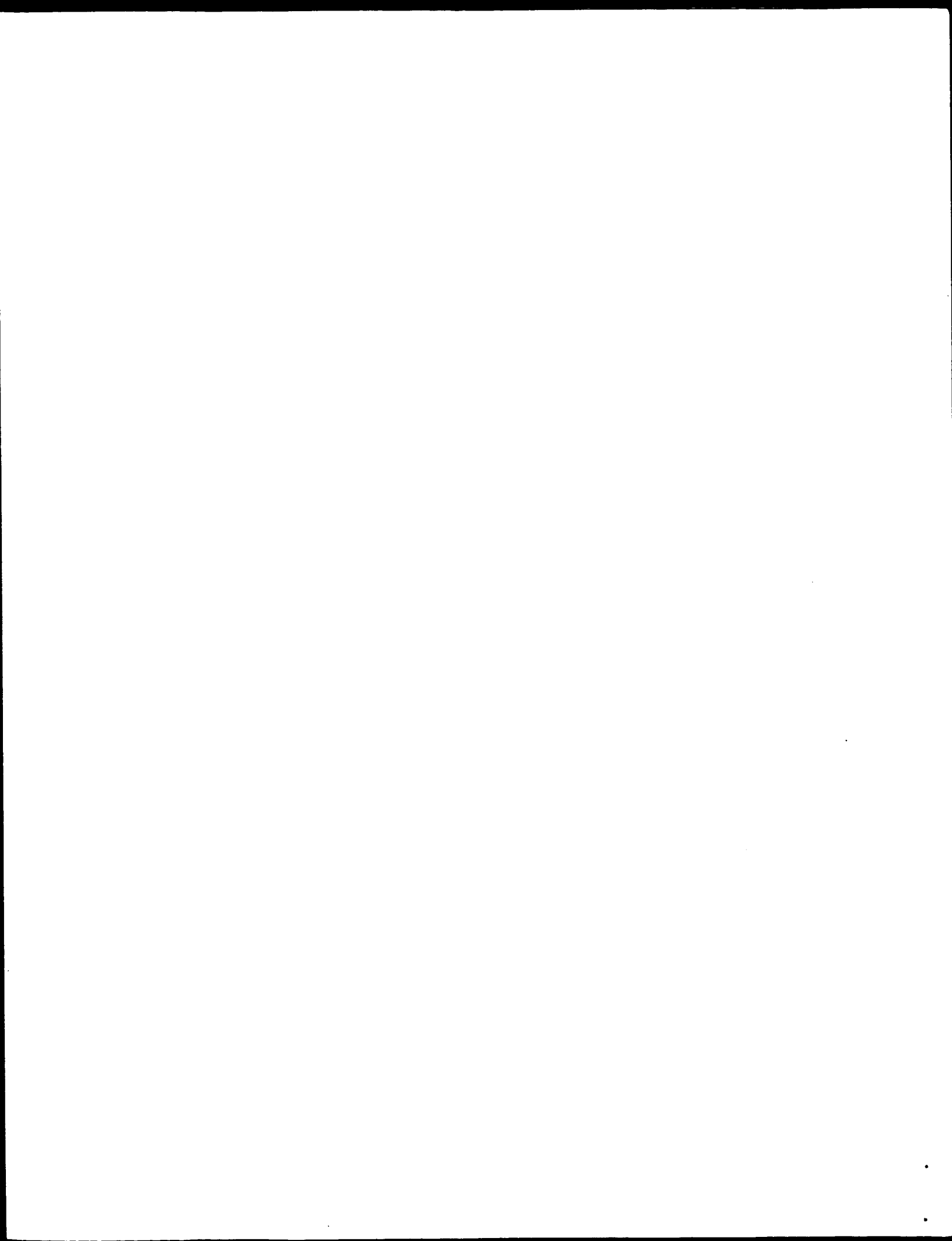
Database sequences: 231628

Database length: 2442594

Search time (sec): 22.730000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

| score_list: | Strd Orig | zscore | EScore | Len | ! Documentation | .. |
|-------------|-----------|--------|--------|-----|-----------------|----|
| Sequence | | | | | | |



OM of: US-09-674-779-4 to: PIR_71.* out_format : pfs

Date: Jul 30, 2002 4:33 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+np.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09674779/runat_30072002_151836_8525/app_query.fasta_1.116
-DB=PIR_71 -QFMT=fastan -SUFFIX=oligo.rpr -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779.ecgn1_1.113 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

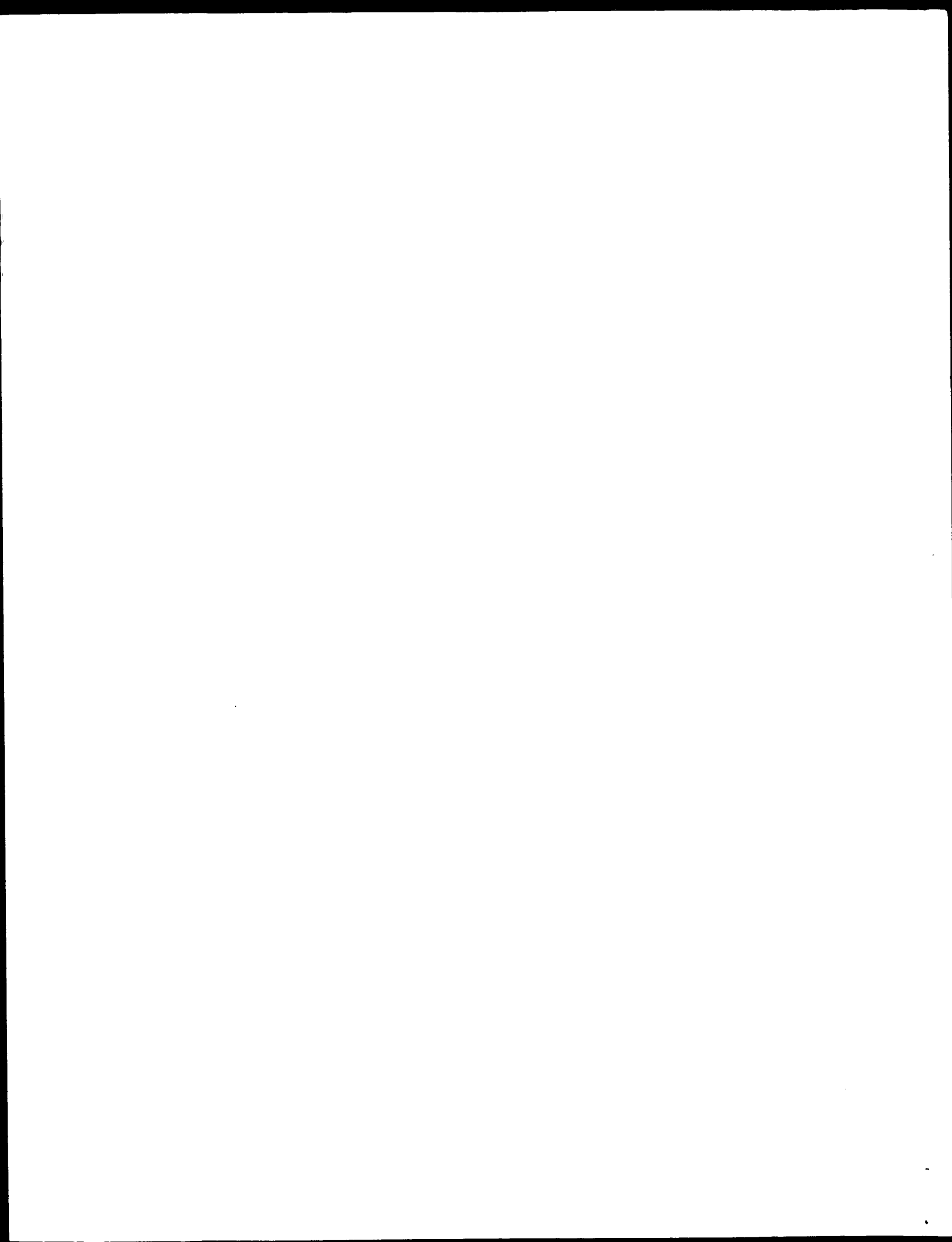
Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: PIR_71.*
Database sequences: 283138
Database length: 96089334
Search time (sec): 28.890000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence Strd Orig zScore EScore Len ! Documentation ..



OM of: US-09-674-779-4 to: SwissProt_40:* out_format : pfs

Date: Jul 30, 2002 4:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09674779/runat_30072002_151837_8635/app_query.fasta_1.116
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=oligo.rsp -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779_@CGN1_1_45 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

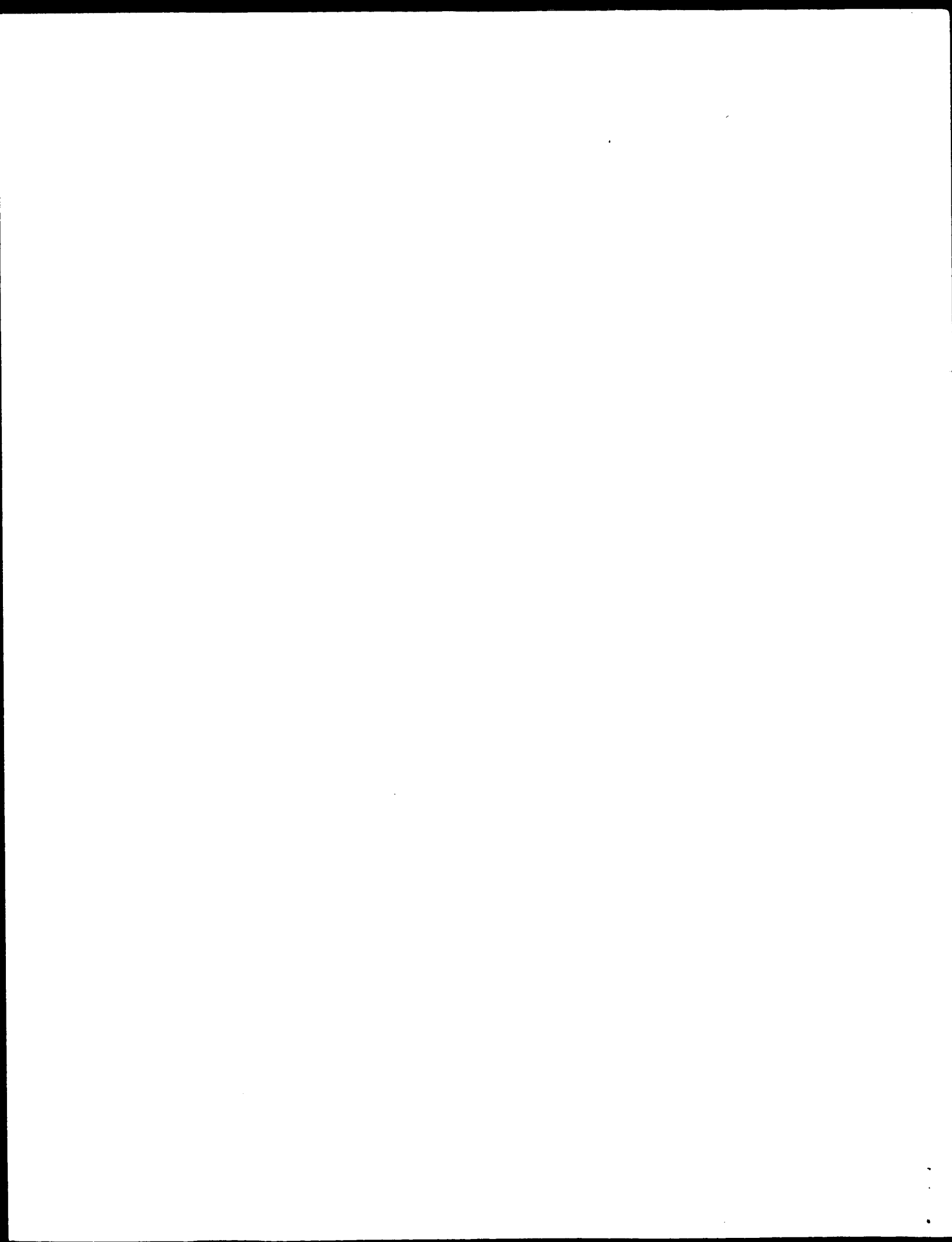
Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 16.530000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

| Sequence | Strd Orig | zScore | EScore | Len | ! Documentation | .. |
|----------|-----------|--------|--------|-----|-----------------|----|
|----------|-----------|--------|--------|-----|-----------------|----|



OM of: US-09-674-779-4 to: SPTREMBL_19:* out_format : pfs

Date: Jul 30, 2002 4:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09674779/runat_30072002_151837_8596/app_query.fasta_1.116
-DB=SPTREMBL_19 -CEMT=fastan -SUFFIX=oligo.rspt -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=15 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGNL_1_205 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-674-779-4
Query length: 60
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 47.120000
```

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

| score_list: | Strd Orig | ZScore | EScore | Len | ! Documentation | .. |
|-------------|-----------|--------|--------|-----|-----------------|----|
| Sequence | | | | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:07:07 ; Search time 29.83 seconds
(without alignments)
930.890 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 1355
Sequence: 1 MKNFNQYFITLISMLVAC.....GAQFSETNSTRHVLKPNKL 250

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1355 | 100.0 | 250 | 22 | AAB60645 |
| 2 | 89.5 | 6.6 | 1377 | 22 | ABG12098 |
| 3 | 88 | 6.5 | 320 | 22 | ABG12098 |
| 4 | 88 | 6.5 | 320 | 22 | ABG12098 |
| 5 | 83.5 | 6.2 | 1218 | 22 | ABG12098 |
| 6 | 83 | 6.1 | 932 | 22 | ABG12098 |
| 7 | 83 | 6.1 | 932 | 22 | ABG12098 |
| 8 | 82.5 | 6.1 | 842 | 22 | ABG12098 |
| 9 | 82 | 6.1 | 308 | 21 | AA74841 |
| 10 | 81 | 6.0 | 382 | 20 | AA737545 |
| 11 | 81 | 6.0 | 1596 | 22 | ABG12098 |

| | | | | | | |
|----|------|-----|------|----|----------|---------------------|
| 12 | 80 | 5.9 | 448 | 22 | AAB73524 | Human transferase |
| 13 | 78 | 5.8 | 233 | 22 | ABG12098 | Novel human diago |
| 14 | 78 | 5.8 | 304 | 21 | AA74840 | Neisseria mening |
| 15 | 78 | 5.8 | 388 | 21 | AA74840 | Neisseria mening |
| 16 | 78 | 5.8 | 578 | 16 | AA75085 | Arabidopsis thalia |
| 17 | 78 | 5.8 | 578 | 16 | AA75085 | Oxalyl-CoA decarbo |
| 18 | 77.5 | 5.7 | 307 | 21 | AA72282 | Human oxalyl-CoA d |
| 19 | 77.5 | 5.7 | 342 | 21 | AA72282 | Arabidopsis thalia |
| 20 | 77.5 | 5.7 | 358 | 21 | AA72282 | Arabidopsis thalia |
| 21 | 77.5 | 5.7 | 2464 | 22 | ABG12098 | Arabidopsis thalia |
| 22 | 77 | 5.7 | 308 | 21 | AA74839 | Drosophila melanog |
| 23 | 77 | 5.7 | 330 | 22 | AA74839 | Neisseria gonorrhoe |
| 24 | 77 | 5.7 | 330 | 22 | AA74839 | Amino acid sequenc |
| 25 | 77 | 5.7 | 330 | 22 | AA74839 | Amino acid sequenc |
| 26 | 77 | 5.7 | 368 | 20 | AA74839 | L. monocytogenes a |
| 27 | 77 | 5.7 | 612 | 22 | ABG12098 | Drosophila melanog |
| 28 | 77 | 5.7 | 935 | 21 | AA75085 | Protein encoded by |
| 29 | 77 | 5.7 | 1783 | 22 | ABG12098 | Drosophila melanog |
| 30 | 76.5 | 5.6 | 867 | 22 | ABG12098 | Novel human diago |
| 31 | 76.5 | 5.6 | 884 | 22 | AA74840 | Human polypeptide |
| 32 | 76.5 | 5.6 | 899 | 22 | AA74840 | Human protein SEQ |
| 33 | 76.5 | 5.6 | 903 | 22 | AA74840 | Human polypeptide |
| 34 | 76.5 | 5.6 | 961 | 22 | AA74840 | Human polypeptide |
| 35 | 76.5 | 5.6 | 1798 | 21 | AA751611 | Human HSGT1 protei |
| 36 | 76 | 5.6 | 352 | 21 | AA74840 | Arabidopsis thalia |
| 37 | 76 | 5.6 | 1139 | 22 | ABG12098 | Drosophila melanog |
| 38 | 75.5 | 5.6 | 300 | 22 | ABG12098 | Drosophila melanog |
| 39 | 75.5 | 5.6 | 3722 | 12 | AA74840 | Cephalosporin anti |
| 40 | 75.5 | 5.6 | 3972 | 21 | AA74840 | S. avermitilis ave |
| 41 | 75.5 | 5.6 | 3972 | 22 | AA74840 | Streptomyces averm |
| 42 | 75.5 | 5.6 | 3972 | 22 | AA74840 | Streptomyces averm |
| 43 | 75 | 5.5 | 306 | 20 | AA74840 | Topoisomerase II b |
| 44 | 75 | 5.5 | 337 | 21 | AA74840 | Human cancer assoc |
| 45 | 75 | 5.5 | 362 | 21 | AA74840 | Human CD40 recepto |

ALIGNMENTS

RESULT 1

ID AAB60645 standard; Protein; 250 AA.

AC AAB60645;

DT 04-MAY-2001 (first entry)

DE Moraxella catarrhalis strain ATCC43617 BASB120 protein.

KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory.

OS Moraxella catarrhalis.

XX WO2001093335-A2.

PN 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EF07361.

XX 03-AUG-1999; 99GB-0018281.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonard J;

DR WPI; 2001-159872/16.

DR N-PSDB; AAF59797.

XX New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
PT strain American Type Culture Collection 43617, for use as therapeutic

PT agents or vaccines against bacterial infections, e.g. otitis media or
 PT pneumonia
 PS Claim 4; Page 64; 75pp; English.
 XX
 XX The invention relates to the Moraxella catarrhalis strain ATCC43617
 CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
 CC invention also relates to immunogenic fragments of the BASB120 protein,
 CC expression vectors and host cells comprising BASB120 nucleic acids, the
 CC recombinant production of BASB120, vaccine compositions comprising the
 CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
 CC compositions comprising the anti-BASB120 antibody, and a method of
 CC identifying a Moraxella catarrhalis infection via the detection of
 CC BASB120 proteins or antibodies. The vaccine compositions of the invention
 CC are useful as prophylactic or therapeutic agents against Moraxella
 CC catarrhalis infections in mammals, particularly humans. Moraxella
 CC catarrhalis is a Gram negative bacterium frequently isolated from the
 CC human upper respiratory tract, which is responsible for several
 CC pathological conditions. It is responsible for about 15% of otitis media
 CC cases in children (which can lead to temporary or permanent hearing
 CC loss). It also causes pneumonia in elderly people, and sinusitis,
 CC nosocomial infections and, less frequently, invasive diseases. BASB120
 CC proteins or nucleotides may additionally be used in screening for novel
 CC antibacterial compounds, and in the diagnosis and staging of infections.
 CC The present sequence represents the Moraxella catarrhalis strain
 CC ATCC43617 BASB120 protein.
 XX
 XX Sequence 250 AA;

Query Match 100.0%; Score 1355; DB 22; Length 250;
 Best Local Similarity 100.0%; Pred. No. 1.7e-145;
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKNFQYFTTTLISSMLVACSAPIPTNPQVSPITKPSVLITKDKIGHDHTHEHDESVSHV 60
 DB 1 mknfqyfttllssmlvacsapipnqpsvptkpsvlltkdkighdhthehdesvshv 60
 QY 61 GLQAHFETWLMQHHATKQEVRYQAYLQSRNLGNYPMPMSQLLTARSQACGHEPYOLPP 120
 DB 61 glqahfetwlmqhhatkqevrvyqaylqsrnlgnypmpsqllttarswqacghepyglpp 120
 QY 121 EHLWGQIVPTLHYDLKSRGTLIPANTOIRSVYRNPELNOCAGGAAMSKHLTNSAIDTW 180
 DB 121 ehllwgqivptlhydlksrgtlipantqirsvyrnpelnocaggaamskhltnsaidtw 180
 QY 181 PDLKTSQALYELQNLRCQYWLHGENONFGLGLYATGAIHLDTCGRKWAQFSETNSI 240
 DB 181 pdleiksqalyelqnlrcqywlhgenqngfllgylatgaihldtqgrkwaqfsetnsi 240
 QY 241 CRHVLPPKNKL 250
 DB 241 crhvlppknkl 250

RESULT 2
 ABB64706
 ID ABB64706 standard; Protein; 1377 AA.
 XX
 AC ABB64706;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 20910.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD

XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL08809.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX
 XX Disclosure; SEQ ID NO 20910; 2lpp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1377 AA;

Query Match 6.6%; Score 89.5; DB 22; Length 1377;
 Best Local Similarity 18.8%; Pred. No. 2.1;
 Matches 51; Conservative 42; Mismatches 87; Indels 91; Gaps 10;
 QY 23 PIPTNQVSPITKPSVLITKDKIGHDHTHEHDESVSHVGLQAHFETWLMQHHATKQ- 78
 DB 292 piknsqpslpkpktpktikstkeksidsaaneslsiv-----vhnatesh 340
 QY 79 -----EVVRYQAYLQ-----SRLGNYPMPMSQLLTARSQACG 112
 DB 341 gscddmetgsgqstsvihhhllqpasahstslridshlsiangrtprkspgin 400
 QY 113 -----HEPYQLPPEHLW-----GQIVPTLHLYQ-----DLKSRGILPA 145
 DB 401 ttdwlmvhrkqpyqvqpthcssttqssidsdsaltpsigdfelksacsvdgskfgiga 460
 QY 146 NTOIRSVYRNPE-----LNOC-----AGGAAMSKHLTNSAIDTWVPLEIK 186
 DB 461 slaprsahkhqnlhssstnlktpicetlvefsssggggpkspfkqsmidpmtlqak 520
 QY 187 SQALYELQNLRCQYWLHGENONFGLGLYAT 217
 DB 521 ttvtssmnl-----qrrgsnhsitlnlhss 547

RESULT 3
 AAG98419
 ID AAG98419 standard; Protein; 320 AA.
 XX
 AC AAG98419;
 XX
 DT 21-SEP-2001 (first entry)
 XX
 DE Escherichia coli protein sequence SEQ ID NO:467.
 XX
 KW Escherichia coli; identification; proliferation; microorganism;
 KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
 KW bacterial growth inhibition.
 XX

OS Escherichia coli.
 PN WO200148209-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 19-DEC-2000; 2000WO-US34419.
 XX
 PR 23-DEC-1999; 99US-0173005.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Forsyth RA, Ohlsen KL, Zyskind JW;
 XX
 DR WPI; 2001-457376/49.
 DR N-PSDB; AAH81475.
 XX
 PT Novel nucleic acids encoding proteins required for Escherichia coli
 PT proliferation, useful for screening for antimicrobial agents -
 XX
 PS Example 4; Page 583; 596pp; English.
 XX
 CC The present invention describes a purified or isolated nucleic acid
 CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
 CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
 CC microorganism is capable of inhibiting proliferation of a microorganism.
 CC (I) have antibacterial and antibiotic activities, and can be used in
 CC gene therapy. Expression of (I) in a microorganism inhibits proliferation
 CC of the microorganism, and the manufactured antibiotic is useful for
 CC reducing the activity or level of a gene product required for
 CC proliferation of a microorganism in a subject, specifically humans. The
 CC nucleic acids that inhibit bacterial growth or proliferation can be used
 CC as antisense therapeutics for killing bacteria. In addition to
 CC therapeutic applications, the nucleic acid sequences complementary to
 CC sequences required for proliferation can be used as diagnostic tools.
 CC For example, nucleic acid probes complementary to proliferation-required
 CC sequences that are specific for particular species of microorganisms can
 CC be used as probes to identify particular microorganism species in
 CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
 CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
 CC represent oligonucleotides, which are used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 320 AA;
 Query Match 6.5%; Score 88; DB 22; Length 320;
 Best Local Similarity 20.8%; Pred. No. 0.35;
 Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;
 QY 8 FITLISMLVACSAPIPTNPQVSPKTPSVLITKDKIGDHTHEHDESVSHVQLQAHFE 67
 Db 10 wltfffaaavalpakantwplp-pagsrlygenkf---hvndgg----- 54
 QY 68 TWLQMHATKQEVVRYOAYLQSLRG--NYLPPMSQLLTARSWQACGHEPYQLPPEHLWG 125
 Db 55 ---sleaakynvgflallqanpgvdypragsvlt----- 89
 QY 126 QIVPTLHLQDLKSRGILPANTQIRSVYRNPELNOCA-----GGAAMSKHLTNSAI 176
 Db 90 --lplqtlldpapgrevivinaealriyyppgknsrvyypigqlggdtlptmtvts 147
 QY 177 D-----IWVPDLKSKQALYELQN-----RLCQY---WLEHGENQ 208
 Db 148 dkrantptwtanir--arykaggtelpavvpagidnpgmhghairlaayggvylhgtna 205
 QY 209 NFGGLG-YATGAIHL 222
 Db 206 dfgimrvssgcirl 220
 RESULT 4
 AAG99078

ID AAG99078 standard; Protein; 320 AA.
 XX
 AC AAG99078;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE E. coli growth and proliferation related protein sequence SEQ ID NO:299.
 XX
 KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
 KW bacterial infection; microorganism.
 XX
 OS Escherichia coli.
 XX
 PN WO200134810-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30950.
 XX
 PR 09-NOV-1999; 99US-0164415.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Forsyth RA, Ohlsen K, Zyskind J;
 XX
 DR WPI; 2001-335933/35.
 DR N-PSDB; AAH84500.
 XX
 PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors
 XX
 PS Claim 19; Page 389-390; 522pp; English.
 XX
 CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG98999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins or other
 CC used to generate reagents and screen small molecule libraries can be
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present
 CC invention.
 XX
 SQ Sequence 320 AA;
 Query Match 6.5%; Score 88; DB 22; Length 320;
 Best Local Similarity 20.8%; Pred. No. 0.35;
 Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;
 QY 8 FITLISMLVACSAPIPTNPQVSPKTPSVLITKDKIGDHTHEHDESVSHVQLQAHFE 67
 Db 10 wltfffaaavalpakantwplp-pagsrlygenkf---hvndgg----- 54
 QY 68 TWLQMHATKQEVVRYOAYLQSLRG--NYLPPMSQLLTARSWQACGHEPYQLPPEHLWG 125
 Db 55 ---sleaakynvgflallqanpgvdypragsvlt----- 89
 QY 126 QIVPTLHLQDLKSRGILPANTQIRSVYRNPELNOCA-----GGAAMSKHLTNSAI 176


```

XX OS Neisseria meningitidis.
XX PN W09957280-A2.
XX PD 11-NOV-1999.
XX PF 30-APR-1999; 99WO-US09346.
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX DR WPI; 2000-062150/05.
XX DR N-PSDB; AA253602.
XX PT Novel Neisserial polypeptides predicted to be useful antigens for
XX PT vaccines and diagnostics -
XX PS Claim 2; Page 636-637; 1453pp; English.
XX AA AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
XX CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides
XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
XX CC PCR primers used in the exemplification of the present invention. The
XX CC polypeptides, the polynucleotides, antibodies and compositions of
XX CC the invention can be used as vaccines, as diagnostic reagents, and as
XX CC immunogenic compositions. The polypeptides can be used in the
XX CC manufacture of medicaments for treating or preventing infection due to
XX CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX CC presence of Neisseria bacteria, or to raise antibodies. They may also
XX CC be used to screen for agonists or antagonists, which may themselves
XX CC have use as antibacterial agents. The polynucleotides of the invention
XX CC may also be used in gene therapy protocols.
XX SQ Sequence 304 AA;

Query Match          5.8%; Score 78; DB 21; Length 304;
Best Local Similarity 21.1%; Pred. No. 4.5;
Matches 41; Conservative 27; Mismatches 74; Indels 52; Gaps 10;

QY 30 VSPITKPSVLITKDKIGDHTHEHDSVSHVGLQAHFE----TWLQ--MHHTKQEVVRY 83
Db 104 iqlk-----aeegghdhhdhdhheghh---hdhgeyqphwvndpvlmsayaqnvaka 154
QY 84 -----QAYLSRLGNLPPMSOL-----LTTARSHWQACGHEPYQ----- 117
Db 155 likadpegkvyqqrlngymqiklksdaqaafnavpaakrkvltghdaifsymgkryhi 214
QY 118 --LPPEHLWGQIVPTLH----LYQDLKSRGILPANTO-----IRSVYR-NPELNQACAGAA 166
Db 215 efiapqgvsseaapsakvaairkregikavftenikdtrmvdriaketgvnvsqkl 274
QY 167 MSKHLTNSADIWV 180
Db 275 ysdalgnapadyi 288

RESULT 15
AAG41671

```

```

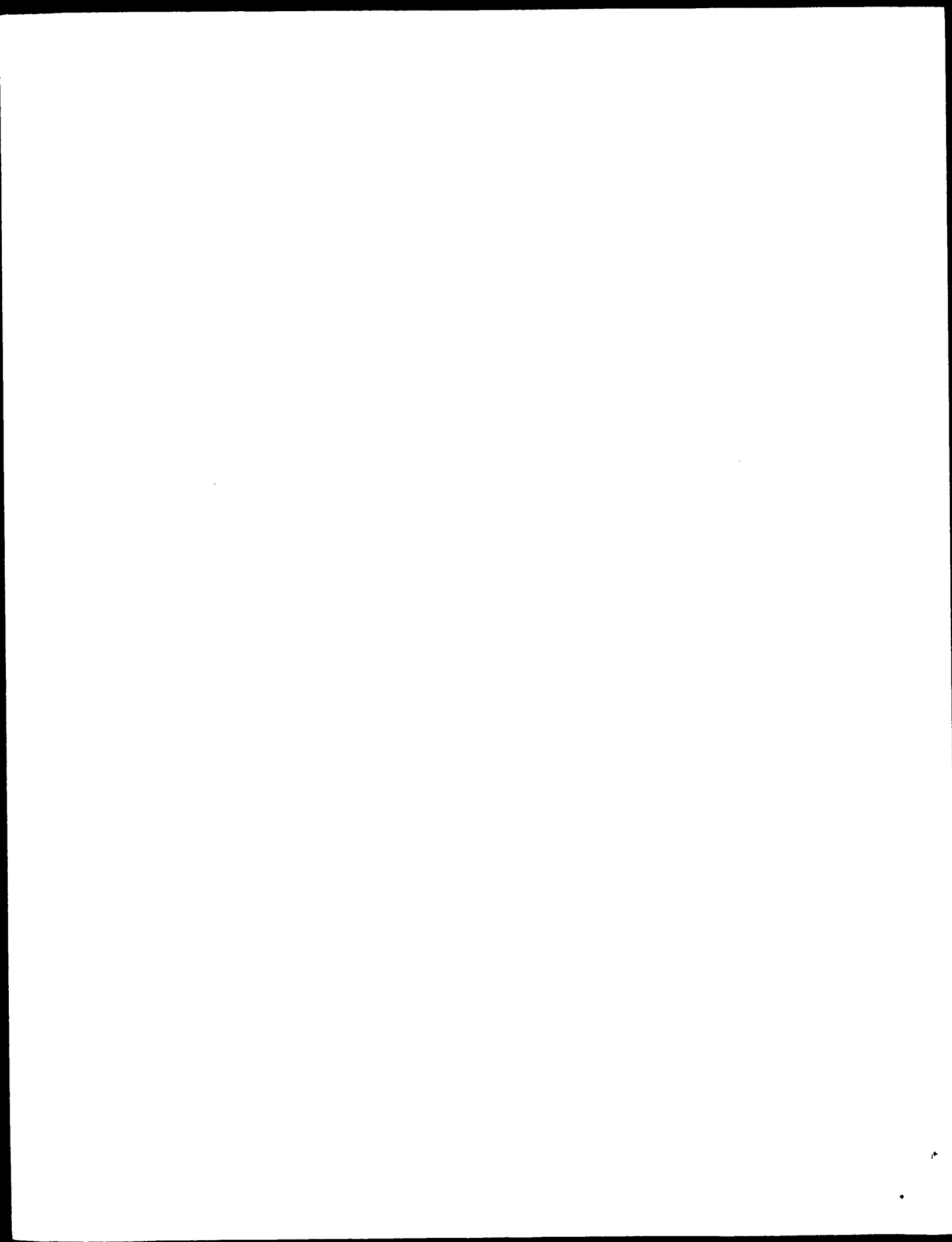
ID XX AAG41671 standard; Protein; 388 AA.
XX AC AAG41671;
XX XX 18-OCT-2000 (first entry)
XX DT
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 51877.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 18-MAY-1999; 99US-0134370.
XX PR 19-MAY-1999; 99US-0134768.
XX PR 20-MAY-1999; 99US-0134941.
XX PR 21-MAY-1999; 99US-0135124.
XX PR 24-MAY-1999; 99US-0135353.
XX PR 25-MAY-1999; 99US-0135629.
XX PR 27-MAY-1999; 99US-0136021.
XX PR 28-MAY-1999; 99US-0136392.
XX PR 01-JUN-1999; 99US-0136782.
XX PR 03-JUN-1999; 99US-0137222.
XX PR 04-JUN-1999; 99US-0137528.
XX PR 07-JUN-1999; 99US-0137502.
XX PR 08-JUN-1999; 99US-0137724.
XX PR 10-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 14-JUN-1999; 99US-0138847.
XX PR 16-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.

```

| | | |
|--|--|---------------|
| PR | 18-JUN-1999; | 99US-0139461. |
| PR | 18-JUN-1999; | 99US-0139462. |
| PR | 18-JUN-1999; | 99US-0139463. |
| PR | 18-JUN-1999; | 99US-0139750. |
| PR | 21-JUN-1999; | 99US-0139763. |
| PR | 21-JUN-1999; | 99US-0139817. |
| PR | 22-JUN-1999; | 99US-0139899. |
| PR | 23-JUN-1999; | 99US-0140353. |
| PR | 23-JUN-1999; | 99US-0140354. |
| PR | 24-JUN-1999; | 99US-0140695. |
| PR | 28-JUN-1999; | 99US-0140823. |
| PR | 30-JUN-1999; | 99US-0140991. |
| PR | 30-JUN-1999; | 99US-0141287. |
| PR | 01-JUL-1999; | 99US-0141842. |
| PR | 01-JUL-1999; | 99US-0142154. |
| PR | 02-JUL-1999; | 99US-0142055. |
| PR | 06-JUL-1999; | 99US-0142390. |
| PR | 08-JUL-1999; | 99US-0142803. |
| PR | 09-JUL-1999; | 99US-0142920. |
| PR | 12-JUL-1999; | 99US-0142977. |
| PR | 13-JUL-1999; | 99US-0143542. |
| PR | 14-JUL-1999; | 99US-0143624. |
| PR | 16-JUL-1999; | 99US-0144005. |
| PR | 16-JUL-1999; | 99US-0144086. |
| PR | 19-JUL-1999; | 99US-0144331. |
| PR | 19-JUL-1999; | 99US-0144332. |
| PR | 19-JUL-1999; | 99US-0144333. |
| PR | 19-JUL-1999; | 99US-0144334. |
| PR | 20-JUL-1999; | 99US-0144335. |
| PR | 20-JUL-1999; | 99US-0144352. |
| PR | 20-JUL-1999; | 99US-0144632. |
| PR | 20-JUL-1999; | 99US-0144884. |
| PR | 21-JUL-1999; | 99US-0144814. |
| PR | 21-JUL-1999; | 99US-0145086. |
| PR | 21-JUL-1999; | 99US-0145088. |
| PR | 22-JUL-1999; | 99US-0145087. |
| PR | 22-JUL-1999; | 99US-0145089. |
| PR | 22-JUL-1999; | 99US-0145192. |
| PR | 23-JUL-1999; | 99US-0145145. |
| PR | 23-JUL-1999; | 99US-0145218. |
| PR | 26-JUL-1999; | 99US-0145276. |
| PR | 27-JUL-1999; | 99US-0145913. |
| PR | 27-JUL-1999; | 99US-0145918. |
| PR | 27-JUL-1999; | 99US-0145919. |
| PR | 28-JUL-1999; | 99US-0145951. |
| PR | 02-AUG-1999; | 99US-0146386. |
| PR | 02-AUG-1999; | 99US-0146388. |
| PR | 03-AUG-1999; | 99US-0146389. |
| PR | 03-AUG-1999; | 99US-0147038. |
| PR | 04-AUG-1999; | 99US-0147204. |
| PR | 05-AUG-1999; | 99US-0147302. |
| PR | 05-AUG-1999; | 99US-0147192. |
| PR | 06-AUG-1999; | 99US-0147260. |
| PR | 06-AUG-1999; | 99US-0147303. |
| PR | 06-AUG-1999; | 99US-0147416. |
| PR | 09-AUG-1999; | 99US-0147493. |
| PR | 09-AUG-1999; | 99US-0147935. |
| PR | 10-AUG-1999; | 99US-0148171. |
| PR | 11-AUG-1999; | 99US-0148319. |
| PR | 12-AUG-1999; | 99US-0148341. |
| PR | 13-AUG-1999; | 99US-0148565. |
| PR | 13-AUG-1999; | 99US-0148684. |
| PR | 16-AUG-1999; | 99US-0149368. |
| PR | 17-AUG-1999; | 99US-0149175. |
| PR | 18-AUG-1999; | 99US-0149426. |
| PR | 20-AUG-1999; | 99US-0149722. |
| PR | 20-AUG-1999; | 99US-0149723. |
| PR | 20-AUG-1999; | 99US-0149929. |
| PR | 23-AUG-1999; | 99US-0149902. |
| PR | 23-AUG-1999; | 99US-0149930. |
| PR | 25-AUG-1999; | 99US-0150566. |
| PR | 26-AUG-1999; | 99US-0150884. |
| PR | 27-AUG-1999; | 99US-0151065. |
| PR | 27-AUG-1999; | 99US-0151066. |
| PR | 27-AUG-1999; | 99US-0151080. |
| PR | 30-AUG-1999; | 99US-0151303. |
| PR | 31-AUG-1999; | 99US-0151438. |
| PR | 01-SEP-1999; | 99US-0151930. |
| PR | 07-SEP-1999; | 99US-0152363. |
| PR | 10-SEP-1999; | 99US-0153070. |
| PR | 13-SEP-1999; | 99US-0153758. |
| PR | 15-SEP-1999; | 99US-0154018. |
| PR | 16-SEP-1999; | 99US-0154039. |
| PR | 20-SEP-1999; | 99US-0154779. |
| PR | 22-SEP-1999; | 99US-0155139. |
| PR | 23-SEP-1999; | 99US-0155486. |
| PR | 24-SEP-1999; | 99US-0155659. |
| PR | 28-SEP-1999; | 99US-0156458. |
| PR | 29-SEP-1999; | 99US-0156596. |
| PR | 04-OCT-1999; | 99US-0157117. |
| PR | 05-OCT-1999; | 99US-0157753. |
| PR | 06-OCT-1999; | 99US-0157865. |
| PR | 07-OCT-1999; | 99US-0158029. |
| PR | 08-OCT-1999; | 99US-0158232. |
| PR | 12-OCT-1999; | 99US-0158369. |
| PR | 13-OCT-1999; | 99US-0159293. |
| PR | 13-OCT-1999; | 99US-0159294. |
| PR | 13-OCT-1999; | 99US-0159295. |
| PR | 14-OCT-1999; | 99US-0159329. |
| PR | 14-OCT-1999; | 99US-0159330. |
| PR | 14-OCT-1999; | 99US-0159331. |
| PR | 14-OCT-1999; | 99US-0159637. |
| PR | 14-OCT-1999; | 99US-0159638. |
| PR | 18-OCT-1999; | 99US-0159584. |
| PR | 21-OCT-1999; | 99US-0160741. |
| PR | 21-OCT-1999; | 99US-0160767. |
| PR | 21-OCT-1999; | 99US-0160768. |
| PR | 21-OCT-1999; | 99US-0160770. |
| PR | 21-OCT-1999; | 99US-0160814. |
| PR | 21-OCT-1999; | 99US-0160815. |
| PR | 22-OCT-1999; | 99US-0160980. |
| PR | 22-OCT-1999; | 99US-0160981. |
| PR | 22-OCT-1999; | 99US-0160989. |
| PR | 25-OCT-1999; | 99US-0161404. |
| PR | 25-OCT-1999; | 99US-0161405. |
| PR | 25-OCT-1999; | 99US-0161406. |
| PR | 26-OCT-1999; | 99US-0161359. |
| PR | 26-OCT-1999; | 99US-0161360. |
| PR | 26-OCT-1999; | 99US-0161361. |
| PR | 28-OCT-1999; | 99US-0161920. |
| PR | 28-OCT-1999; | 99US-0161992. |
| PR | 28-OCT-1999; | 99US-0161993. |
| PR | 29-OCT-1999; | 99US-0162142. |
| Query Match 5.8%; Score 78; DB 21; Length 388; | | |
| Best Local Similarity 21.8%; Pred. No. 6.5; | | |
| Matches 65; Conservative 32; Mismatches 87; Indels 114; Gaps 16; | | |
| QY | 31 SPIKTP-----SVLITKDKTIGDHTHEHDES-----VSHVGLQAHFETWLOMH-HAT 76 | |
| Db | 13 ssllkkp1livlvcitvsvllvitymvpqnnhskssacaglsrgcqaalsgwlphvhrkf 72 | |
| QY | 77 KOEVRYQAYLQSLRGLNVLPPMSQLLTARSWQACGH-EPYQLPPEHLWGOIVPTLHLYQ 135 | |
| Db | 73 tdeevaarvikkil--rlpp-----alcakskiafmfltpgtlpfeklwdk-----ffq 120 | |
| QY | 136 DLKSR-GILPANTQIRSVYRNPNLQOCAGGAMSKHLTNSAIDIWVPDLKISQ-----188 | |
| Db | 121 qcqgrfsiyhpsrlrphv-----isrhfs-----dreihsdhvtwgr 158 | |
| QY | 189 -ALYELONRLOCOYWLEHGENQNF-----GLGLY 215 | |

Db 159 ismvdarrillanalepdnqhfvllsesciplhtfdytyrylmhanvsvfidsfedlgph 218
QY 216 ATGAIHLD-----TQGRK---W-----GAQFSETNSICRHVLPKNK 249
Db 219 gtgr-hmdhmlpeiprgdfrkgacqwtmkrqhavivmadglyyskfreycrpgveank 275

Search completed: July 30, 2002, 16:15:08
Job time: 481 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:07:52 ; Search time 13.04 Seconds
(without alignments)
468.282 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 1355
Sequence: 1 MNFNQYFTITLISMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|----|--------------------|
| 1 | 78 | 5.8 | 578 | 1 | US-08-458-120-2 |
| 2 | 78 | 5.8 | 578 | 2 | US-08-867-970-2 |
| 3 | 78 | 5.8 | 578 | 4 | US-09-326-217-2 |
| 4 | 77 | 5.7 | 368 | 3 | US-08-972-902-2 |
| 5 | 75 | 5.5 | 755 | 3 | US-09-071-101-2 |
| 6 | 75 | 5.5 | 755 | 3 | US-09-369-618-2 |
| 7 | 75 | 5.5 | 755 | 3 | US-09-369-617-2 |
| 8 | 74 | 5.5 | 462 | 4 | US-09-165-241-1 |
| 9 | 73 | 5.4 | 420 | 2 | US-08-588-2588-40 |
| 10 | 73 | 5.4 | 420 | 5 | PCT-US96-08285-40 |
| 11 | 73 | 5.4 | 1141 | 1 | US-08-363-300-2 |
| 12 | 71.5 | 5.3 | 140 | 4 | US-09-439-313-477 |
| 13 | 71 | 5.2 | 1091 | 3 | US-08-986-485-5 |
| 14 | 70.5 | 5.2 | 305 | 1 | US-08-844-485-6 |
| 15 | 70 | 5.2 | 437 | 4 | US-09-353-332-2 |
| 16 | 70 | 5.2 | 512 | 4 | US-09-149-922-2 |
| 17 | 69 | 5.1 | 410 | 6 | 5177197-1 |
| 18 | 69 | 5.1 | 1394 | 6 | 5177197-30 |
| 19 | 68.5 | 5.1 | 436 | 3 | US-08-660-347-2 |
| 20 | 68.5 | 5.1 | 473 | 2 | US-08-700-670A-23 |
| 21 | 68.5 | 5.1 | 883 | 2 | US-08-596-366-2 |
| 22 | 68.5 | 5.1 | 883 | 2 | US-08-967-104-2 |
| 23 | 68 | 5.0 | 111 | 3 | US-08-875-811-22 |
| 24 | 68 | 5.0 | 226 | 4 | US-08-858-207A-417 |
| 25 | 68 | 5.0 | 473 | 4 | US-09-286-691-15 |
| 26 | 68 | 5.0 | 473 | 4 | US-09-687-147-15 |
| 27 | 68 | 5.0 | 657 | 3 | US-08-893-852A-3 |

| | | | | | | |
|----|------|-----|------|---|-------------------|--------------------|
| 28 | 68 | 5.0 | 657 | 4 | US-08-821-818-3 | Sequence 3, Appli |
| 29 | 68 | 5.0 | 4472 | 2 | US-08-804-227C-2 | Sequence 2, Appli |
| 30 | 67.5 | 5.0 | 622 | 2 | US-08-459-146-2 | Sequence 2, Appli |
| 31 | 67.5 | 5.0 | 622 | 2 | US-08-459-065-2 | Sequence 2, Appli |
| 32 | 67.5 | 5.0 | 712 | 2 | US-08-468-5768-17 | Sequence 17, Appli |
| 33 | 67.5 | 5.0 | 712 | 2 | US-08-468-579B-17 | Sequence 17, Appli |
| 34 | 67.5 | 5.0 | 712 | 3 | US-08-468-577B-17 | Sequence 17, Appli |
| 35 | 67.5 | 5.0 | 820 | 4 | US-09-173-914-2 | Sequence 2, Appli |
| 36 | 67.5 | 5.0 | 1245 | 2 | US-08-801-263A-3 | Sequence 3, Appli |
| 37 | 67.5 | 5.0 | 1245 | 3 | US-08-801-263A-6 | Sequence 6, Appli |
| 38 | 67.5 | 5.0 | 1245 | 3 | US-09-102-248-3 | Sequence 3, Appli |
| 39 | 67.5 | 5.0 | 1245 | 3 | US-09-102-248-6 | Sequence 6, Appli |
| 40 | 67.5 | 5.0 | 2485 | 5 | PCT-US94-00198-1 | Sequence 1, Appli |
| 41 | 67.5 | 5.0 | 2485 | 5 | PCT-US94-00198-2 | Sequence 2, Appli |
| 42 | 67.5 | 5.0 | 2818 | 1 | US-08-510-284-1 | Sequence 1, Appli |
| 43 | 67.5 | 5.0 | 2818 | 1 | US-08-411-389-2 | Sequence 2, Appli |
| 44 | 67.5 | 5.0 | 2818 | 2 | US-08-449-933-2 | Sequence 2, Appli |
| 45 | 67.5 | 5.0 | 2818 | 4 | US-07-966-049A-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-08-458-120-2
; Sequence 2, Application US/08458120
; Patent No. 5635616
; GENERAL INFORMATION:
; APPLICANT: OLSEN, ET AL.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,120
; FILING DATE: SUBMITTED HEREWITH
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05561
; FILING DATE: 18 MAY 94
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-458-120-2

Query Match 5.8%; Score 78; DB 1; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 71; Indels 58; Gaps 11;

QY 82 RYQ---AYLQ-----SRILNYLPPMSQLL-----TTARSMQACGHEPVQLPPEHLWQIV 128

Db 294 RYQPDVKFIQVDICAEELGNNVKKPAVTLGNHVTAKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176
Db 352 -TLREKMKSEAAASKELASKSLPMNYTTFVYHVQEQLPDRDCFVWSEGA-----NTM 402
QY 177 DIWVDPLEIKSQAELYELQNLRCQWLEHGENQNFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNLPRHRLDAGTGTGTGVLGFAIAAAVAVAKDRSPGOWIICV 452
QY 177 DIWVDPLEIKSQAELYELQNLRCQWLEHGENQNFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNLPRHRLDAGTGTGTGVLGFAIAAAVAVAKDRSPGOWIICV 452
QY 225 QGFRKMGWQAFSETNSICRHVLP 246
Db 453 EGDSAFGSGMEVETICRYNLP 474

RESULT 2
US-08-867-970-2
; Sequence 2, Application US/08867970
; Patent No. 5945273
; GENERAL INFORMATION:
; APPLICANT: Olsen, Henrik S.
; APPLICANT: Coleman, Timothy
; APPLICANT: Adams, Mark D.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,970
; FILING DATE: JUN-3-97
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/458,120
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/05561
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF119D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-867-970-2

Query Match 5.8%; Score 78; DB 2; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 24; Mismatches 71; Indels 58; Gaps 11;
QY 82 RYQ---AYLQ-----SRIGNVLPMSQL-----TTARSWQACGHEPYQLPPEHLGQIV 128
Db 294 RYQPDVKFIQVDICAEELGNNVKKPAVTLGNHVTAKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176

Db 352 -TLREKMKSEAAASKELASKSLPMNYTTFVYHVQEQLPDRDCFVWSEGA-----NTM 402
QY 177 DIWVDPLEIKSQAELYELQNLRCQWLEHGENQNFGLGL---YATGAIHLD-----T 224
Db 403 DI-----GRTVLQNLPRHRLDAGTGTGTGVLGFAIAAAVAVAKDRSPGOWIICV 452
QY 225 QGFRKMGWQAFSETNSICRHVLP 246
Db 453 EGDSAFGSGMEVETICRYNLP 474

RESULT 3
US-09-326-217-2
; Sequence 2, Application US/09326217
; Patent No. 6200796
; GENERAL INFORMATION:
; APPLICANT: Olsen, Henrik S.
; APPLICANT: Coleman, Timothy
; APPLICANT: Adams, Mark D.
; TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/326,217
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,970
; FILING DATE: JUN-03-97
; APPLICATION NUMBER: US 08/458,120
; FILING DATE: 02-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/05561
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF119D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-326-217-2

Query Match 5.8%; Score 78; DB 4; Length 578;
Best Local Similarity 24.3%; Pred. No. 3.6;
Matches 49; Conservative 24; Mismatches 71; Indels 58; Gaps 11;
QY 82 RYQ---AYLQ-----SRIGNVLPMSQL-----TTARSWQACGHEPYQLPPEHLGQIV 128
Db 294 RYQPDVKFIQVDICAEELGNNVKKPAVTLGNHVTAKOLLELDKTPWQYPPESKWK-- 351
QY 129 PTLH-----LYQDLKSRGILPANTQIRSVYRNPEL-NQC---AGGAAMSKHLTNSAI 176
Db 352 -TLREKMKSEAAASKELASKSLPMNYTTFVYHVQEQLPDRDCFVWSEGA-----NTM 402

QY 177 DIWVPLEIKSQALYELQNLRCQYWLHGENQNFGLGL--YATGAIHLDT-----T 224
 Db 403 DI-----GRTVLQNYLPRHRLDAGTFTGTMVGLGFATAAAVAVKDRSPGOWIICV 452
 QY 225 QGFRKWAQFSETNSTRVLP 246
 Db 453 EGDSAFGFGMEVETICRYNLP 474

RESULT 4
 US-08-972-902-2
 ; Sequence 2, Application US/08972902
 ; Patent No. 6099848
 ; GENERAL INFORMATION:
 ; APPLICANT: Frankel, Fred R.
 ; APPLICANT: Portnoy, Daniel A.
 ; TITLE OF INVENTION: BACTERIAL VACCINE VECTOR AND METHODS OF
 ; USE THEREOF
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
 ; STREET: One Commerce Square, 2005 Market Street, 22nd
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: U.S.A.
 ; ZIP: 19103-7086
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/972.902
 ; FILING DATE: 18-NOV-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Doyle-Leary, Kathryn
 ; REGISTRATION NUMBER: 36,317
 ; REFERENCE/DOCKET NUMBER: 9596-51
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-965-1284
 ; TELEFAX: 215-567-2991
 ; TELEX: 831-494
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 368 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-972-902-2

Query Match 5.7%; Score 77; DB 3; Length 368;
 Best Local Similarity 21.3%; Pred. No. 2.3;
 Matches 46; Conservative 34; Mismatches 98; Indels 38; Gaps 9;

QY 50 THEHDESV---SHVQLQAHFETLQMHATKQEVVRYQAYLQSRGLNGYLPMPMSQLTT-- 104
 Db 90 TRKEDANLAANKHISLTVFEDWLE--NLTLATLRLHLKVDSCMGL-----GIRTEE 142
 QY 105 ARSQWACGHEPYQLPPEHLWGQIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNOGAG 164
 Db 143 ARREATSTNDHQLQLEGIVTHFAATQLETSYFEQQLAKFQILTSLKKRPTVYHTANS 202
 QY 165 AA--MSKHLTNSAIDIW-----PDLEIKSQALYELQNLRCQYW-----LEHGENON 209
 Db 203 AASLLQIQIGDAIRFGISMYGLTPSTEIKTSLPFELKPALALYTEMVHVHVELAPGDSVS 262
 QY 210 FGLGLYATGAIHLDTQGRKWA--QFSETNSTRH 243

Db 263 YGATYATATE-----REWVATLPIGYADGLIRH 289

RESULT 5
 US-09-071-101-2
 ; Sequence 2, Application US/09071101
 ; Patent No. 6013503
 ; GENERAL INFORMATION:
 ; APPLICANT: Lok, Si
 ; APPLICANT: Jaspers, Stephen R.
 ; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics, Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,101
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, Gary E.
 ; REGISTRATION NUMBER: 31,648
 ; REFERENCE/DOCKET NUMBER: 97-05
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-442-6673
 ; TELEFAX: 206-442-6678
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 755 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; FRAGMENT TYPE: internal
 ; US-09-071-101-2

Query Match 5.5%; Score 75; DB 3; Length 755;
 Best Local Similarity 22.0%; Pred. No. 12;
 Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;

QY 14 SSMVLVACASAPIPTNPOVSPIKTPSVLITKDKIGDHH--THEHDESVSHVGLQA----- 64
 Db 340 STLTITTYSSGVATDPOI-----VTTDL---HHGCTDQHTGTSASAPLAAGMALAL 387
 QY 65 ---HFTWLQMHATKQEVVRYQ--AYLOS-----RLGNYLPMPMSQLTTA 105
 Db 388 EANPFLTWRDMQHL-----VVRASKFAHLQADWRTNGVRQVSHHYGYGLLDAGLLVDTA 443
 QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYQDLKSR--GILPANTQIRSVYRNPELNOGAG 163
 Db 444 RTW-----LPTQ-----PQKCAVRVQSRPTLPL-----IYIRENVVSACAG 481
 QY 164 GRAMSKHLTNSAIDIWDPDLEIKSQALYELQNLRCQYWLHGE-----NQNFGLG-----LYA 216
 Db 482 -----LHNST-----RSLEHVQALTLTSYRRGDLTSLTSPMGTRSLVA 522
 QY 217 TGAIHLDTOGFRKW 230
 Db 523 IRPLDVSTEGYNNW 536

Query Match 5.5%; Score 75; DB 3; Length 755;
 Best Local Similarity 22.0%; Pred. No. 12;
 Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;

QY 14 SSMVLACSAPIPTNPQVSPKTPSVLITKDKIGDHH--THEHDESVSHVGLQA-----64
 Db 340 STLTITVSSGVATDPOI-----VTTDL---HHGCTDQHTGTSASAPLAAGMIALAL 387

QY 65 ---HFETWLOMHATKQEVVRQ--AYLOS-----RLGNVLPMSOLLTTA 105
 Db 388 EAMPFLTWDRMQL-----VVRASKPAHLQAEDWRTNGVRQVSHHYGGLDAGLLVDTA 443

QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYODLKS--GILPANTQIRSVYRNPELNQCAG 163
 Db 444 RTW-----LPTQ-----PQRKCAVRVOSRPTPLPL-----IYIRENVSAACAG 481

QY 164 GAAMSKHLTNSAIDIWVPDLKTSQALYELQNLCOYWLEHGE---NQNFGLG---LYA 216
 Db 482 -----LHNSI-----RSLEHVQAQLTSLYSRRGDLSEISLTSPMGTRSTLVA 522

QY 217 TGAHLDTOGFRKW 230
 Db 523 IRPLDVSTEGYNW 536

RESULT 8

US-09-165-241-1
 ; Sequence 1, Application US/09165241
 ; Patent No. 6130077
 ; GENERAL INFORMATION:
 ; APPLICANT: Henry Yue
 ; APPLICANT: Karl J. Guegler
 ; APPLICANT: Neil C. Corley
 ; APPLICANT: Janice Au-Young
 ; TITLE OF INVENTION: HUMAN CYTOCHROME P450
 ; FILE REFERENCE: EP-0608 US
 ; CURRENT APPLICATION NUMBER: US/09/165,241
 ; CURRENT FILING DATE: 1998-10-01
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1
 ; LENGTH: 462
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE: -
 ; OTHER INFORMATION: 991729
 US-09-165-241-1

Query Match 5.5%; Score 74; DB 4; Length 462;
 Best Local Similarity 25.2%; Pred. No. 7.2;
 Matches 35; Conservative 21; Mismatches 67; Indels 16; Gaps 7;

QY 113 HEPYOLPPEHLW--GOIVPTLHLYODLK-----SRGILPANTQIRSVYRNPELNQCAGGA 165
 Db 62 HERVG-PVVSFWFGRRLLVSLGTVDLVKQHLNPKTLDPETMLKSLR-----YQSGGGS 116

QY 166 AMSKHLTNSAIDIWVPD--LEIKSQALYELQNLCOYWLEHGENQFGLGYATG-AIHL 223
 Db 117 VSENHMRKKLYENGVTDSLKSNFALLKLSEBLLDKWLSYPETQHVPLSQHMLGFAMKSV 176

QY 224 TQGRKWAQFSETNSICR 242
 Db 177 TQ--VMVGSTFEDDOEVIR 193

RESULT 9

US-08-588-258B-40
 ; Sequence 40, Application US/08588258B
 ; Patent No. 5929207
 ; GENERAL INFORMATION:

US-09-369-618-2
 ; Sequence 2, Application US/09369618
 ; Patent No. 6100041
 ; GENERAL INFORMATION:
 ; APPLICANT: Lok, Si
 ; APPLICANT: Jaspers, Stephen R.
 ; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
 ; FILE REFERENCE: 97-05D2
 ; CURRENT APPLICATION NUMBER: US/09/369,618
 ; CURRENT FILING DATE: 1999-08-06
 ; EARLIER APPLICATION NUMBER: US 09/071,101
 ; EARLIER FILING DATE: 1998-05-01
 ; EARLIER APPLICATION NUMBER: US 60/044,015
 ; EARLIER FILING DATE: 1998-05-06
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 755
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-369-618-2

Query Match 5.5%; Score 75; DB 3; Length 755;
 Best Local Similarity 22.0%; Pred. No. 12;
 Matches 56; Conservative 32; Mismatches 72; Indels 94; Gaps 15;

QY 14 SSMVLACSAPIPTNPQVSPKTPSVLITKDKIGDHH--THEHDESVSHVGLQA-----64
 Db 340 STLTITVSSGVATDPOI-----VTTDL---HHGCTDQHTGTSASAPLAAGMIALAL 387

QY 65 ---HFETWLOMHATKQEVVRQ--AYLOS-----RLGNVLPMSOLLTTA 105
 Db 388 EAMPFLTWDRMQL-----VVRASKPAHLQAEDWRTNGVRQVSHHYGGLDAGLLVDTA 443

QY 106 RSWQACGHEPYQLPPEHLWGQIVPTLHLYODLKS--GILPANTQIRSVYRNPELNQCAG 163
 Db 444 RTW-----LPTQ-----PQRKCAVRVOSRPTPLPL-----IYIRENVSAACAG 481

QY 164 GAAMSKHLTNSAIDIWVPDLKTSQALYELQNLCOYWLEHGE---NQNFGLG---LYA 216
 Db 482 -----LHNSI-----RSLEHVQAQLTSLYSRRGDLSEISLTSPMGTRSTLVA 522

QY 217 TGAHLDTOGFRKW 230
 Db 523 IRPLDVSTEGYNW 536

RESULT 7

US-09-369-617-2
 ; Sequence 2, Application US/09369617
 ; Patent No. 6127162
 ; GENERAL INFORMATION:
 ; APPLICANT: Lok, Si
 ; APPLICANT: Jaspers, Stephen R.
 ; TITLE OF INVENTION: HUMAN PROHORMONE CONVERTASE 4
 ; FILE REFERENCE: 97-05D1
 ; CURRENT APPLICATION NUMBER: US/09/369,617
 ; CURRENT FILING DATE: 1999-08-06
 ; EARLIER APPLICATION NUMBER: US 09/071,101
 ; EARLIER FILING DATE: 1998-05-01
 ; EARLIER APPLICATION NUMBER: US 60/044,015
 ; EARLIER FILING DATE: 1998-05-06
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 755
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-369-617-2

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG3019
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-587 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA44573.1; PID:g1742190; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3763
A:Map position: linear chromosome

Query Match 7.5%; Score 101; DB 2; Length 587;
Best Local Similarity 22.5%; Pred. No. 0.64;
Matches 38; Conservative 31; Mismatches 58; Indels 42; Gaps 9;
QY 74 HATKQEVVRYQAYLQSRIGNY----LPPMSQLLTARSQACGHEPYQLPPEHLWGQIVP 129
DB 28 HTREKAVITFK-----RNGKYDQKGLQELNREL---RDWRR--NQPTMDPR----- 69
QY 130 TLHLYQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLE 184
DB 70 LFDLVWEVYRRSGATDYINVVSAPRSPTENGLLRTTKGVAEKSQHMLGKAMDFYIPG-- 127
QY 185 IKSQALYELQNLRCQYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
DB 128 VKLATLREIGMQM-----QIGGVGYPTSGSPFVHMDVGGVRAW 166

RESULT 3
B98265
hypothetical protein AGR_L_2143 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
R:Accession: B98265
A:Authors: B. Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorillo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 234, 2323-2328, 2001
A:Title: Genome sequence of the plant pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: B98265
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-634 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89644.1; PID:g15159543; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2143
A:Map position: linear chromosome

Query Match 7.5%; Score 101; DB 2; Length 634;
Best Local Similarity 22.5%; Pred. No. 0.71; Mismatches 58; Indels 42; Gaps 9;
Matches 38; Conservative 31; Mismatches 58; Indels 42; Gaps 9;
QY 74 HATKQEVVRYQAYLQSRIGNY----LPPMSQLLTARSQACGHEPYQLPPEHLWGQIVP 129
DB 75 HTREKAVITFK-----RNGKYDQKGLQELNREL---RDWRR--NQPTMDPR----- 116
QY 130 TLHLYQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLE 184
DB 117 LFDLVWEVYRRSGATDYINVVSAPRSPTENGLLRTTKGVAEKSQHMLGKAMDFYIPG-- 174
QY 185 IKSQALYELQNLRCQYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
DB 175 VKLATLREIGMQM-----QIGGVGYPTSGSPFVHMDVGGVRAW 213

RESULT 4
AH3510
hypothetical membrane associated protein BMEII0010 [imported] - Brucella melitensis (str

C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AH3510
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanov
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AH3510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-637 <KUR>
A:Cross-references: GB:AE008918; PIDN:AA53251.1; PID:g17984130; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEII0010
A:Map position: II

Query Match 7.1%; Score 96.5; DB 2; Length 637;
Best Local Similarity 22.9%; Pred. No. 1.8;
Matches 40; Conservative 31; Mismatches 53; Indels 51; Gaps 12;
QY 72 MHATKQEVVRYQAYLQSRIGNYLP--MSQLLTARSQACGHEPYQLP--EHLWGQI 127
DB 37 VHTGEKAEI---AFKED--GRFLPDGLKRLNVLRLDWR--NEPTRMDPRLEFDLIW--- 85
QY 128 VPTLHLYQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDP 182
DB 86 ----QVQSTGSRVYI---TVVSAYRSPATNMLRSTRGVAKKSOHMLGRAMDYIPG 137
QY 183 LEI-KSQAL---YELQNLRCQYWLHGENQNFGLGLYATGA---IHLDTQGRFKW 230
DB 138 VPLAKLRAIGMRYQIG-----GVGYPRSGSPFVHMDVGNVRHW 176

RESULT 5
AH0644
probable exported protein STY1255 [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AH0644
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
., S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AH0644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08339.1; PID:g16502384; GSPDB:GN00176
C:Genetics:
A:Gene: STY1255

Query Match 6.9%; Score 93; DB 2; Length 263;
Best Local Similarity 19.8%; Pred. No. 1.2;
Matches 56; Conservative 37; Mismatches 92; Indels 98; Gaps 13;
QY 9 ITTLSSMLVACSAPIPTNPQVSPKTP-SVLITKDKIGDHTHEDSVSHVGLQAHFE 67
DB 13 LLALATATMLA---LPAQANTWPLPPGSRVGENKF---HVVEDDGG----- 54
QY 68 TWLQMHATKQEVRYQAYLQSRIG--NVLPPMSQLLTARSQACGHEPYQLPPEHLWG 125
DB 55 ----SLATAKYNVGFLLQANPGIDPYVPAGSVLT----- 89
QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176

Db 90 --IPLQTLDPAPREGIVINLAELRLYYQPKNTVTYPIGIGLQGGDTLTPTMTTIS 147

QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY---WLEHGENQ 208

Db 148 DKRANPTWTPTNIR--ARYKAQGIIDLPVAVPAGDPNMGHHAIRLAAYGGVYLLHGTNA 205

QY 209 NFGGLG-YATGAIHLDTQGRKWAQFSETNSICRHVLPKKNL 250

Db 206 GFGIGMRVSSGCIRL-----RGGDIETLFRQVTPGTRKV 238

RESULT

AD2215

Two-component hybrid sensor and regulator all3275 [imported] - Anabaena sp. (strain PCC

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AD2215

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2215

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-889 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074974.1; PID:g17132370; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all3275

Query Match

Best Local Similarity 6.6%; Score 90; DB 2; Length 889;

Matches 45; Conservative 32; Mismatches 78; Indels 68; Gaps 10;

QY 15 SMLVACSAPITPNQVSPDKTSPVLTIKDKI--GDHHTH-----EHDESIVSHVGLQA 64

Db 300 SLLVQCAVSLFVGTAITERSIOVELAVEKVRGEYQTRAELSEKLIQLNESLVAI---- 355

QY 65 HFETLQMHATKQEVRYQAYLQSRGLNVLPPMSQ--LLTARSWA--CGH----- 113

Db 356 --NDCLESREKDELQRLQSRGLNVLPPMSQ--LLTARSWA--CGH----- 113

QY 114 EPVOLPPEHLWGQIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNQACAGAAAMSKHLIN 173

Db 414 EPEDLICKNYEQWTRT-----KCTDFEREYRRS----- 442

QY 174 SAIDWIWPDLEIKSQALYELQNLCQWLE-HGENONFGLGLY 215

Db 443 -----LAENIPVHFELYELWD-----MWLEVHAYPSDGLGIF 476

RESULT

F64855

Ycfs protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999

C:Accession: F64855

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64855

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-320 <BLAT>

A:Cross-references: GB:AE000211; GB:U00096; NID:g1787345; PIDN:AAC74197.1; PID:g1787356;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: ycfs

C:Superfamily: conserved hypothetical protein b0819

C:Keywords: periplasmic space

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-320/Product: ycfs protein #status predicted <MAT>

Query Match 6.5%; Score 88; DB 2; Length 320;

Best Local Similarity 20.8%; Pred. No. 4.1;

Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTILSSMLVACSAPITPNQVSPDKTSPVLTIKDKIGDHTHEDSHVSHVGLQAHPF 67

Db 10 WLUTFTTFAAVALALPAKANTWPLP--PAGSRLVGENKF---HVVDGG----- 54

QY 68 TWLQMHATKQEVRYQAYLQSRGLG--NYLPPMSQLLTARSQACGHEPYQLPPEHLWG 125

Db 55 ---SLEAIAKYNVGFALLQANPGVDVYPRAGSVLT----- 89

QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176

Db 90 --IPLQTLDPAPREGIVINIAELRLYYPPGKNSVTYPIGIGLQGGDTLTPTMTTIS 147

QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY---WLEHGENQ 208

Db 148 DKRANPTWTPTNIR--ARYKAQGIELPAPVAVPAGLDNPMGHHAIRLAAYGGVYLLHGTNA 205

QY 209 NFGGLG-YATGAIHL 222

Db 206 DFGIGMRVSSGCIRL 220

RESULT

8

C90815

hypothetical protein Ecs1491 [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: C90815

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90815

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-320 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034914.1; PID:g13360955; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs1491

C:Superfamily: conserved hypothetical protein b0819

Query Match

Best Local Similarity 6.5%; Score 88; DB 2; Length 320;

Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTILSSMLVACSAPITPNQVSPDKTSPVLTIKDKIGDHTHEDSHVSHVGLQAHPF 67

Db 10 WLUTFTTFAAVALALPAKANTWPLP--PAGSRLVGENKF---HVVDGG----- 54

QY 68 TWLQMHATKQEVRYQAYLQSRGLG--NYLPPMSQLLTARSQACGHEPYQLPPEHLWG 125

Db 55 ---SLEAIAKYNVGFALLQANPGVDVYPRAGSVLT----- 89

QY 126 QIVPTLHLYQDLKSRGILPANTQIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176

Db 90 --IPLQTLDPAPREGIVINIAELRLYYPPGKNSVTYPIGIGLQGGDTLTPTMTTIS 147

QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY---WLEHGENQ 208

Db 148 DKRANPTWTPTNIR--ARYKAQGIELPAPVAVPAGLDNPMGHHAIRLAAYGGVYLLHGTNA 205

QY 209 NFGLGL-YATGAHL 222
 Db 206 DFGIGMRVSSGCIRL 220

RESULT 9

G85674
 hypothetical protein ycf5 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85674
 R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 K.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamou, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85674
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <STO>
 A:Cross-references: GB:AE005174; NID:g12514663; PIDN:AAG55859.1; GSPDB:GN00145; UWGP:Z17
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ycf5
 C:Superfamily: conserved hypothetical protein b0819

Query Match 6.5%; Score 88; DB 2; Length 320;

Best Local Similarity 20.8%; Pred. No. 4.1;
 Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITTSSMLVACSAPIPNQVSPKTPSVLITKDKIGDHTTHEDESVHVLQAHE 67
 Db 10 WLTFFFAAVALAPAKANTWPLP-PAGSRVGENKF---HVENDGG----- 54
 QY 68 TWLOMHATKQEVVRYQVLYQSLG--NVLPPMSQLTTTARSWQAGGHPYQLPPEHLWG 125
 Db 55 ---SLEAIKKYVNGFLALQANPGVDYVPRAGSVLT----- 89
 QY 126 QIVPTLHLVODLKSRLPANTQIRSVYRNPELNOC-----GGAAMSKHLTNSAI 176
 Db 90 --IPLQTLDPDREGIVINIAELRLYYPGKNSVTVPIGIGQLGGDTLPTWTWTVS 147
 QY 177 D-----IWVPLEIKSQALYELQN-----RLCOY---WLEHGENQ 208
 Db 148 DKRANPTWTPTANIR--ARYKAQGIELPAVVPAGPDNPMGHAIRLAAYGGVYLLHGTNA 205
 QY 209 NFGLGL-YATGAHL 222
 Db 206 DFGIGMRVSSGCIRL 220

RESULT 10

D71490
 probable exodeoxyribonuclease V, gamma - Chlamydia trachomatis (serotype D, strain UW3/C)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: D71490
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
 A:Reference number: A71570; MUID:99000809
 A:Accession: D71490
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1004 <ARN>
 A:Cross-references: GB:AE001334; GB:AE001273; NID:g3329078; PIDN:AAC68244.1; PID:g332908
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: recC

Query Match 6.5%; Score 88; DB 2; Length 1004;

Best Local Similarity 19.7%; Pred. No. 17;
 Matches 52; Conservative 40; Mismatches 88; Indels 84; Gaps 11;

QY 17 LVACSAPIPNQVSPKTPSVLITKDKIGDHTTHE 52
 Db 754 LLSKTPPTNTYLSAFTESLYTDVQDSVSKRLETQKQDPATTPFVSVDQLFHDPLHP 813
 QY 53 HDESVS-----HVGLOAHFETWLMQMHATKQEVVR 82
 Db 814 NDQOVSPLLSLPLKNTLHQTTHGVCVSGVYLFMSHPG-EAEKKT--OKTHGFPPKDAFE 870
 QY 83 YOAYLQSRGLNYLPPMSQLL-TTARSWQACGHE-----PYQLPPPEHLWGQIVPTLHLIQ 135
 Db 871 LESYLSYLALQASHLLPKKATILRVTPHDIEPLTPPFSSPESYL-----IRAIHLYE 926
 QY 136 DLKSRGI-LPANTQIRSVYRNPELNOCAGGAAMSKHLTNSAID-----IWVPLD---EI 185
 Db 927 LLQQAIPPLPSAOWEYIKTKTDSASQCI-----KKLLDSEEDPLTSSFWFHNRTTEEI 980
 QY 186 KSQALYELQNLRCQYWLHGENQN 209
 Db 981 CSRLSNDVLSQLSLFINQDSQQN 1004

RESULT 11

A11295
 phosphoribosylformylglycinamide synthetase I [imported] - Listeria monocytogenes (s
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: A11295
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
 Ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 ; Title: Comparative genomics of Listeria species;
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: A11295
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-739 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC99847.1; PID:g16411223; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: purQ
 C:Superfamily: phosphoribosylformylglycinamide synthase component II

Query Match 6.5%; Score 87.5; DB 2; Length 739;

Best Local Similarity 19.5%; Pred. No. 13;
 Matches 48; Conservative 42; Mismatches 83; Indels 73; Gaps 13;

QY 40 ITKDKIGDHTTHEDESVHVLQAHEFETWLMQMHATKQEVVRYQVLYQSRGLNYLPPMS 99
 Db 352 VTDDKM--YKLIHGEVAVNPVDALEAD-APVYHKPSKEPTRYQAFQESEA--FVPMAD 406
 QY 100 QLLTTARSWQAGGHPYQLPPEHLWGQIVPTLHLIQDLKSRGILP----- 144
 Db 407 DVVGV---WKELLAQPTIASKRHIVEQ-----YDYQVTDVAVVPGSDAAIVRVRGTEKA 458
 QY 145 ----ANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVPLEIKSQ 188
 Db 459 TAMTDCNSRYLYLDPPEVGAIAVAEARNIVCSGGKPLA---ITDGLNFCNPE---KPE 512
 QY 189 ALVELQ---NRLCOVWLE-----HGE-----NQFGLGLYATGAI-----HLDLQ 225
 Db 513 IFWEIEKAADGISACACLEDLTVPVLSGNVSLYNETDGTGIYPTPVIGMWGLVEDLAHITQ 572
 QY 226 GPRKWG 231
 Db 573 DFKNSG 578

Db 128 LLVAPLAGISINPPGSKFTTATNVYTPQALGYSRDSDSPA----- 166
 QY 76 TKQEVVRVQAYLQS--RLGNLYP-----PMSQLLTARSQWACG-HEPYQLPPPEHLW 124
 Db 169 -----RYQGREVYSRLRLTYFSPSTLAYQVNDLSVGLSVGFHQVALNEDFRAPNQ--- 219
 QY 125 GQIVPTLHYLDKSRGILPANTQIRSVYRNPELNOACAGGAAKSHLTNSAID----- 177
 Db 220 -----LLGLLOQTKEIGCLPGMOEILEVF-----FNVCGNIGTPFONLANIDLDMQOOLSP 270
 QY 178 -----IWPDLKISQALYELQNL---COYWLHGEN-QNFGGLGYAT--GAI 220
 Db 271 SPNLGVLWEPTDWFANGATYQSESRMLKGYRDYDGGQGWFTGVHKSLLGAI 325
 RESULT 14
 E36792
 hypothetical protein ORF58 - ictalurid herpesvirus 1 (strain auburn 1)
 C:Species: ictalurid herpesvirus 1
 A:Note: host Ictalurus punctatus (channel catfish)
 C:Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 24-Nov-1999
 C:Accession: E36792
 R: Davison, A.J.
 submitted to GenBank, January 1992
 A:Description: Channel catfish virus: a new type of herpesvirus.
 A:Reference number: A36804
 A:Accession: E36792
 A:Molecule type: DNA
 A:Residues: 1-599 <DAV>
 A:Cross-references: GB:M75136; NID:g331209; PIDN:AAA88161.1; PID:g331268
 R: Davison, A.J.
 Virology 186, 9-14, 1992
 A:Title: Channel catfish virus: a new type of herpesvirus.
 A:Reference number: A39447; MUID:92087490
 A:Contents: annotation
 A:Note: neither protein nor nucleic acid sequence is given
 C:Genetics: 58
 A:Gene: 58
 C:Superfamily: ictalurid herpesvirus 1 hypothetical protein ORF58
 Query Match 6.2%; Score 84; DB 2; Length 599;
 Best Local Similarity 18.2%; Pred. No. 20;
 Matches 58; Conservative 50; Mismatches 95; Indels 116; Gaps
 QY 1 MKNFNQVFTTLISSML-----VACSAPIPTNPQ--VSPIKTPSVLITKDKIGDHHTH 51
 Db 267 MAMFRYPIDRLRLRMIMDYFGSLGKCLATFTPPKYTVSGERSHSIVIKNHVDKHYLN 326
 QY 52 EHDESVSHVGLQAHFEFWLQMHATKQEVRYQAYLQSRGLGNVLPMSQLLTARSQWAC 111
 Db 327 DNSFELAHVMDRAMPPTCYIH-----DDIDVKAMIQSVISRVVRMIQDTQVRLQELSA 381
 QY 112 GHPEY-----QLPPE-----HLWGQT-VPTL-----HLYQDL--KSRG 141
 Db 382 GNKLHFHFFNQLPPPEMNNLDTSKYNPLVDHALQGGQGVPLQVGNPEQIYTDWITEMTA 441
 QY 142 ILPA-----NTQIRS-----VYRN-----PELNOACGAAM----- 167
 Db 442 LLPRIGHVMEFMEWSQELSSKTLCLLLPKIARDLVNPLVTLSDDDTGNAMLIFAHKI 501
 QY 168 -----SKHLTNSAIDTWVPLETKSOALYELQN 195
 Db 502 VIATAMLHILHGENLETHDLGPQTGTGVGVGEPKCHTGAOLREWFIDLO-KTIAL-DAPS 559
 QY 196 RLC-----QYWLHGENQNPF 210
 Db 560 THCAGCADFWLTHGSDPNF 578
 RESULT 15
 B82162
 hypothetical protein VC1750 [imported] - Vibrio cholerae (strain N16961 sero

Search completed: July 30, 2002, 16:15:54
Job time: 447 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:15:13 ; Search time 11.07 seconds
(without alignments)
874.425 Million cell updates/sec

Title: US-09-674-779-2

Perfect score: 1355

Sequence: 1 MKNFNYFITLSSMLVAC.....GAQFSETNCRHVLPKNKL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 88 | 6.5 | 320 | 1 YCFS_ECOLI | P75954 escherichia |
| 2 | 84 | 6.2 | 599 | 1 VG58_HSV11 | Q00157 ictaluriid h |
| 3 | 83 | 6.1 | 209 | 1 L0LB_HFEIN | P45270 haemophilus |
| 4 | 83 | 6.1 | 528 | 1 YAC8_YEAST | P39734 saccharomyc |
| 5 | 82.5 | 6.1 | 297 | 1 SYGA_BACHD | Q9kd49 bacillus ha |
| 6 | 82.5 | 6.1 | 365 | 1 MLTA_ECOLI | P46885 escherichia |
| 7 | 80.5 | 5.9 | 474 | 1 PEDA_LACHE | Q48558 lactobacill |
| 8 | 80 | 5.9 | 369 | 1 CD5S_MOUSE | Q35926 mus musculu |
| 9 | 77.5 | 5.7 | 594 | 1 NIR_SPIOL | P05314 spinacia ol |
| 10 | 77 | 5.7 | 350 | 1 REDD_STRCO | P16922 streptomyce |
| 11 | 77 | 5.7 | 506 | 1 VLL_BPVA | P08341 bovine papi |
| 12 | 77 | 5.7 | 1107 | 1 ALA2_ARATH | P98205 arabidopsis |
| 13 | 76.5 | 5.6 | 316 | 1 COAA_BACHD | Q9kx77 bacillus ha |
| 14 | 76.5 | 5.6 | 852 | 1 RBMA_RAT | P70501 rattus norv |
| 15 | 76 | 5.6 | 865 | 1 MUTS_BACHD | Q9kac0 bacillus ha |
| 16 | 75.5 | 5.6 | 390 | 1 VASS_BPGA | P07394 bacteriopho |
| 17 | 75.5 | 5.6 | 563 | 1 PHNL_DESFR | P18188 desulfovibr |
| 18 | 75 | 5.5 | 321 | 1 CATO_HUMAN | P43234 homo sapien |
| 19 | 75 | 5.5 | 368 | 1 ALR_LISMO | O85045 listeria mo |
| 20 | 75 | 5.5 | 746 | 1 YUY1_CAEEL | P54073 caenorhabdi |
| 21 | 74.5 | 5.5 | 467 | 1 VNSS_TSWVL | P26003 tomato spot |
| 22 | 74.5 | 5.5 | 912 | 1 KPCM_HUMAN | Q15139 homo sapien |
| 23 | 74 | 5.5 | 438 | 1 XYLE_THEYO | Q9kqu2 thermoanaer |
| 24 | 74 | 5.5 | 1053 | 1 CAPP_SYNP6 | P06516 synectococc |
| 25 | 74 | 5.5 | 1189 | 1 PTNE_MOUSE | Q62130 mus musculu |
| 26 | 74 | 5.5 | 2190 | 1 CCAD_CHICK | O73700 gallus gall |
| 27 | 74 | 5.5 | 2210 | 1 RRPL_EBOSM | Q66802 ebola virus |
| 28 | 73.5 | 5.4 | 449 | 1 TRB2_AERPE | Q9y9h2 aeropyrum p |
| 29 | 73.5 | 5.4 | 595 | 1 YE72_HUMAN | Q9nx95 homo sapien |
| 30 | 73.5 | 5.4 | 615 | 1 CPO_DROME | Q01617 drosophila |
| 31 | 73.5 | 5.4 | 1440 | 1 Y8EO_PSEAE | P14403 j genome po |
| 32 | 73 | 5.4 | 212 | 1 Y8EO_PSEAE | Q9hvt1 pseudomonas |
| 33 | 73 | 5.4 | 448 | 1 XYLE_BACLI | P77832 bacillus li |

34 73 5.4 477 1 RGS7_RAT
35 73 5.4 495 1 RGS7_HUMAN
36 73 5.4 1237 1 POL4_DROME
37 73 5.4 1273 1 YAR2_SCHPO
38 73 5.4 2594 1 7LES_DROVI
39 73 5.4 3674 1 SPCR_HUMAN
40 72.5 5.4 470 1 VL2_HPV54
41 72.5 5.4 872 1 VP2_ROTFC
42 72.5 5.4 975 1 PMPA_CHLTR
43 72.5 5.4 3421 1 TEGU_HSVB
44 72 5.3 283 1 YG20_YEAST
45 72 5.3 393 1 IDH_STRMU

P49803 rattus norv
P49802 homo sapien
P10394 drosophila
Q10135 schizosacch
P20806 drosophila
Q9nrc6 homo sapien
Q81023 human papil
P26191 porcine rot
O84417 chlamydia t
P28955 equine herp
P53068 saccharomyc
Q59940 streptococc

ALIGNMENTS

RESULT 1
YCFS_ECOLI
ID YCFS_ECOLI STANDARD; PRT; 320 AA.
AC P75954;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycfs precursor.
GN YCFS OR B1113
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE ERPF/YBIS/YCFS/YNHG FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL: AE000211; AAC74197.1; -
EMBL: D90746; BAA35928.1; -
EMBL: D90747; BAA35933.1; -
EcoGene: EGI3437; ycfs.
InterPro: IPR002482; LysM.
Pfam: PF01476; LysM; 1
SMART: SM00257; LysM; 1
Hypothetical protein; Periplasmic; Signal; Complete proteome.
SYNOPSIS
SIGNAL 1 23
POTENTIAL.

FT CHAIN 24 320 HYPOTHETICAL PROTEIN YCFS.
SQ SEQUENCE 320 AA; 34636 MW; 89378A8DFD60359B CRC64;

Query Match
Best Local Similarity 6.5%; Score 88; DB 1; Length 320;
Matches 53; Conservative 30; Mismatches 88; Indels 84; Gaps 11;

QY 8 FITLSSMLVACSAPIPNQVSPKTPSLTKDKIGDHTHEDES VSHVGLQAHE 67
DB 10 WLTFTEFAAVALPAKANTWPLP-PAGSRVGENKF--HVVDGG----- 54
QY 68 TWLQMHATKQEVRYQAVYLSRLG--NYLPPMSQLLTARSQACGHEPYOLPEHLWG 125
DB 55 ---SLEAIKKYVGFALLQANPGVDVYVPRAGSVLT----- 89
QY 126 QYPTTHLYODLSKRGILPANTOIRSVYRNPELNQCA-----GGAAMSKHLTNSAI 176
DB 90 --IPLQTLPLDPAREGINIAELRYYPYPPGKNSVTYPIGIGQLGGDTLTFTVMTVVS 147
QY 177 D-----IWPVDEIKSQAALYELQN-----RLCQY---WLEHGENQ 208
DB 148 DKRANPTWTFTANIR-ARYKAQGIELPVAVPAGLDNPMGHHAIRLAAVGGVYLLHGTNA 205
QY 209 NFGGL-LYATGAHL 222
DB 206 DFGIGMRVSSGCIRL 220

RESULT 2
VG58_HSVII STANDARD; PRT; 599 AA.
AC Q00157;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Hypothetical gene 58 protein.
GN 58.
OS Itatulid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Unclassified Herpesviridae.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AUBURN 1;
RC MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus."
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M75136; AAA88161.1; .
DR PIR; E36792; E36792.
KW Hypothetical protein.
SQ SEQUENCE 599 AA; 67476 MW; 991C2B524F0F6BEF CRC64;

Query Match
Best Local Similarity 6.2%; Score 84; DB 1; Length 599;
Matches 58; Conservative 50; Mismatches 95; Indels 116; Gaps 15;

QY 1 MKNFNQVFTTLISML-----VACSAPIPNQV--VSPKTPSVLITKDKIGDHTH 51
DB 267 MAMFRYPIDRLRLRMIMDYFGSGLKCTLATFTPPKYTVSGERSRHSIVIKNHVDKHYLN 326
QY 52 EHDES VSHVGLQAHEFTWLMQHHATKQEVRYQAVYLSRLGSLNGLVLPMSQLLTARSQAC 111

DB 327 DNSKFLAHVMDRAMPPTCYIH-----DDIDVKAMIQSVISRVVRMIQDTQVRLQELSA 381
QY 112 GHEPY-----OLPPE-----HLWGQI-VPTL-----HLYQDL--KSRG 141
DB 382 GNKLFFHFFNOLPPEMMNLDISFKYNPLVDHALQOGKGVPLQYGNPEQIYTDIMETMA 441
QY 142 ILPA-----NTQIRS-----VYRN-----PELNOCAGGAAM----- 167
DB 442 LLPRIGHMVEPMESQELSSSKTCLLLPKIARDLNPVPLVTLSDNDDTGNAMLIJFAH 501
QY 168 -----SKHLNSAIDIWVPLEIKSQAALYELQN 195
DB 502 YIATAMLIHLHGENLETHDLGPQTGTGVGPEPKHKTGAQLREWFIDLQ-KTIAL-DAPS 559
QY 196 RLC-----QYWLEHGENQNF 210
DB 560 THCACGADFWLTHGSDPNF 578

RESULT 3
LOLB_HAEIN STANDARD; PRT; 209 AA.
AC P45270;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer-membrane lipoprotein lolB precursor.
GN LOLB OR HEMM OR H1607.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D.; Adams M.D.; White O.; Clayton R.A.; Kirkness E.F.;
RA Klenzwe A.R.; Bult C.J.; Tomb J.-F.; Dougherty B.A.; Merrick J.M.;
RA McKenney K.; Sutton G.; Fitzhugh W.; Fields C.A.; Gocayne J.D.;
RA Scott J.D.; Shirley R.; Liu L.-I.; Glodek A.; Kelley J.M.;
RA Weidman J.F.; Phillips C.A.; Spriggs T.; Hedblom E.; Cotton M.D.;
RA Utterback T.R.; Hanna M.C.; Nguyen D.T.; Saudek D.M.; Brandon R.C.;
RA Fine L.D.; Fritchman J.L.; Fuhrmann J.L.; Geoghagen N.S.M.;
RA Gnehm C.L.; McDonald L.A.; Small K.V.; Fraser C.M.; Smith H.O.;
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: PLAYS A CRITICAL ROLE IN THE INCORPORATION OF
CC LIPOPROTEINS IN THE OUTER MEMBRANE AFTER THEY ARE RELEASED BY
CC THE LOLA PROTEIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (By similarity).
CC -!- SIMILARITY: BELONGS TO THE LOLB FAMILY.
CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE INVOLVED IN DELTA-
CC AMINOLEVULINIC ACID BIOSYNTHESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32834; AAC23251.1; .
DR TIGR; H1607; .
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Chaperone; Outer membrane; Lipoprotein; Transport; Protein transport;
KW Signal; Complete proteome.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 209 OUTER-MEMBRANE LIPOPROTEIN LOLB.

FT LIPID 22 22 N-ACYL DIGLYCERIDE (BY SIMILARITY).
SQ SEQUENCE 209 AA; 24193 MW; 0555F02F13E852A1 CRC64;

Query Match 6.1%; Score 83; DB 1; Length 209;
Best Local Similarity 23.0%; Pred. No. 1.7;
Matches 59; Conservative 36; Mismatches 87; Indels 74; Gaps 17;

QY 1 MKNFQYFITTLLISSMLVACSAPI--PNNQVSPKTPSVLITKDKIGDHTHEHDSVS 58
Db 4 KMTF-KFTALFAITLDCMERPTNVQYID-KTDAI----- 41

QY 59 HVGLOAHFETWLOMHATKQVRYQAYLQSRGLNGLPMSOLLTTARSQACGHEPYQL 118
Db 42 -----W-QQHLQIKIQSYQA--KGQIG-YISP-TERSSRFQW-----YQN 80

QY 119 PPEHLWGOIVTLHLQDLKSRGILPANTQIRSVYRNPPELNOQAGGAAKSHLTNSAIDI 178
Db 81 PKSY-----TLKLSLISKSTLWQHQSGMTISDNGNQ--QSAANSKLILQEIIGM 131

QY 179 WVPDLTKSQALYELQNLRCQYWL--EHEGNOFGLGL-VATGAHLDTGFRKQWQAQ-- 233
Db 132 DVP-----LEH--LAYWLKQGPAMNADYQVGTNHLGAFTHVDG--SOWTADYL 177

QY 234 -FSETNSICRHVLPKN 248
Db 178 TYHSNNSMPENILKN 193

RESULT 4
YAC8_YEAST STANDARD; PRT; 528 AA.
AC P39734; P39733;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 58.8 kDa protein in MTO4-DRS2 intergenic region.
GN YAL029W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
cerevisiae";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN [2]
RP REVISIONS.
RA Vo D.T.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U12980; AAC05004.1; -
DR SGD; S0000026; YAL028W.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 308 328 POLY-SER.
FT TRANSMEM 505 521 POTENTIAL.
FT DOMAIN 512 517 POLY-ILE.
SQ SEQUENCE 528 AA; 58750 MW; E5A3CC7C6D60977A CRC64;

Query Match 6.1%; Score 83; DB 1; Length 528;
Best Local Similarity 19.4%; Pred. No. 5.2;
Matches 41; Conservative 35; Mismatches 89; Indels 45; Gaps 6;

QY 3 NFQYFITTLLISSMLVACSAPIPTNPQVSPKTPSVLITKDKIGDHTHEHDSVSHVGL 62
Db 308 SFSQSSSSSSSSSSSSSTFSQSQVAVDPLEPPGNTYSSSSNLNSLSDLDYQYQRHGL 367

QY 63 QAHFETWLOMHATKQVRYQAYLQSRGLNY--LPPMSOLLTTARSQACGHEPYQL- 118
Db 368 QLQ-QTEALLKSLKDEVLDKNDLVKNIANFDKIVKELDRSRFTGKVELVEDYLMN 426

QY 119 -----PPEHLWGOIVTPLHL-----YQDLKSRGILPANTQIRSVYRNPPELNOQAGGAA 166
Db 427 LKODFDKENPESPEARLSDTINTNAKLQDLEKR-----MASCKDRLA 469

QY 167 MSKHLTNSAIDIWVPDLTKSQALYELQNLRL 197
Db 470 SRKEVMR-----KMESLLSLENSL 488

RESULT 5
SYGA_BACHD STANDARD; PRT; 297 AA.
AC Q9KD49;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glycyl-tRNA synthetase alpha chain (EC 6.1.1.14) (Glycine--tRNA ligase
alpha chain) (GLYRS).
GN GLYO OR BH1370.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AP001511; BAB05089.1; -
DR InterPro; IPR002106; AA_trna_ligase_II.
DR InterPro; IPR002310; trna_synt_2e.
DR Pfam; PF02091; trna_synt_2e; 1.
DR PRINTS; PR01044; TRNASYNTHCA.
DR PROSITE; PS00179; AA-trna_ligase_II_1; FALSE_NEG.
DR PROSITE; PS00339; AA-trna_ligase_II_2; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 297 AA; 34420 MW; F27D6C2C74969E7 CRC64;

```

Query Match          6.1%; Score 82.5; DB 1; Length 297;
Best Local Similarity 24.3%; Pred. No. 2.8; 56; Indels 59; Gaps 12;
Matches 44; Conservative 22; Mismatches 56; Indels 59; Gaps 12;

QY 67 ETWLOMHATKQEVRYQAYLQSLGNLYLPPMSQLLTTA-RSQWACGHEPYQLPPEHLWG 125
DB 12 EYW-----SKQNCILLQAYDEKAGTMSPTMLRTTIGPEPNVAYVEPSRRPADRGY 65
QY 126 -----QIV-----PT--LHLYQD-LKSRGILPANTOIRSV---YRNPQLNQACAG 163
DB 66 ENPNRLYQHQRFPKMPSTNIQIQLYLDLRLGALGNPLEHDIRVDENWENPSLG-CAG 124
QY 164 GAAMSKHLNTSAIDWVPLEI-----KSOALYELQNLRCQYWLHGE 206
DB 125 -----LQWEWLDGNEITQFTYFOQVGLNPLNPSAIIYCLE-RLASY-IDQKE 172
QY 207 N 207
DB 173 N 173

RESULT 6
MLTA_ECOLI STANDARD; PRT; 365 AA.
AC P46885; P76638; Q46928;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Membrane-bound lytic murein transglycosylase A precursor (EC 3.2.1.-)
DE (Murein hydrolase A) (Mlt38).
GN MLTA OR MLT OR B2813 OR Z4130 OR ECS3673.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX STRAIN=K12;
RX MEDLINE=97431497; PubMed=9287002;
RA Lomatzsch J., Templin M.F., Kraft A.R., Vollmer W., Hoeltje J.-V.;
RT "Outer membrane localization of murein hydrolases: MltA, a third
RT lipoprotein lytic transglycosylase in Escherichia coli.";
RL J. Bacteriol. 179:5465-5470(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

```

```

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [5]
RP SEQUENCE OF 147-161; 204-213 AND 258-280, AND CHARACTERIZATION.
RX MEDLINE=94117367; PubMed=8288527;
RA Ursinus A., Hoeltje J.-V.;
RT "Purification and properties of a membrane-bound lytic
RT transglycosylase from Escherichia coli.";
RL J. Bacteriol. 176:338-343(1994).
RN [6]
RP INTERACTION WITH MIPA AND MRCP/PONB.
RX STRAIN=ATCC 53338 / MC1061;
RX MEDLINE=99156961; PubMed=10037771;
RA Vollmer W., von Rechenberg M., Hoeltje J.-V.;
RT "Demonstration of molecular interactions between the murein polymerase
RT PBP1B, the lytic transglycosylase MltA, and the scaffolding protein
RT MipA of Escherichia coli.";
RL J. Biol. Chem. 274:6726-6734(1999).
CC -!- FUNCTION: MUREIN-DEGRADING ENZYME. MAY PLAY A ROLE IN RECYCLING OF
CC MUROPEPTIDES DURING CELL ELONGATION AND/OR CELL DIVISION. OPTIMAL
CC ACTIVITY IS BETWEEN PH 4.0 AND 4.5; LOSES ITS ACTIVITY RAPIDLY AT
CC TEMPERATURES ABOVE 30 DEGREES CELSIUS. DEGRADES MUREIN GLYCAN
CC STRANDS AND INSOLUBLE, HIGH-MOLECULAR WEIGHT MUREIN SACCULI.
CC -!- CATALYTIC ACTIVITY: CLEAVAGE OF THE BETA-1,4-GLYCOSIDIC BOND
CC BETWEEN N-ACETYLGLUCOSAMINE AND N-ACETYLGLUCOSAMINE RESIDUES,
CC THEREBY CONSERVING THE ENERGY IN A NEWLY SYNTHESIZED
CC 1,6-ANHYDROBOND IN THE MURAMIC ACID RESIDUE.
CC -!- SUBUNIT: FORMS A TRIMERIC COMPLEX WITH MRCP/PONB AND MIPA IN
CC VITRO.
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32224; AAC45723.1; -
CC EMBL; AE0000365; AAC75855.1; -
CC EMBL; U29581; AAB40463.1; ALT_INIT.
CC EMBL; AE005509; AAG57927.1; -
CC EMBL; AP002563; BAB37096.1; -
CC EcoGene; EG13085; mltA.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC Cell wall; Hydrolase; Glycosidase; Signal; Lipoprotein;
CC Outer membrane; Multigene family; Complete proteome.
CC SIGNAL 1 20 PROBABLE
CC CHAIN 21 365 MEMBRANE-BOUND LYTIC MUREIN
CC TRANSGLYCOSYLASE A.
CC LIPID 21 21 N-ACYL DIGLYCERIDE (PROBABLE).
CC CONFLICT 346 346 H -> N (IN REF. 1).
CC SEQUENCE 365 AA; 40410 MW; 5ECB92C1E8D5969 CRC64;

Query Match          6.1%; Score 82.5; DB 1; Length 365;
Best Local Similarity 19.1%; Pred. No. 3.7;
Matches 57; Conservative 43; Mismatches 88; Indels 111; Gaps 16;

QY 6 QYFITLSSMLVACSA-PIPTNPQVSPK--TPSVLTK-DKIG-DHHTHEDESVSHV 60
DB 7 KYLLMGTVVAMIAACSSKPTDRGQYKDGKFTQPSFLVQNPDAVGAPINAGDFAEQINHI 66
QY 61 -----GLOAH-----FETWL-----QMHATKQEVRYQAYLQSLR 91
DB 67 RNSSPRLYGNSQNVNAVQEWLNRAGDTRNMRQFGIDAWQMEGADNYGNVQFTGY----- 121
QY 92 GNYLPPMSOLLTARSQWACGHEPYQLPPEHLWGQIVFTLHLYQDLKSRGILPANTQIRS 151

```

Db 122 --YTPVQIARHTROGEFQ---YPIYRMPK-----RGLPSRAEIIYA 158
QY 152 VYRNPELNOAGGAMSKHL---INSAIDIWVDPLEIKSQALYELQNLRCQYWEHGENQ 208
Db 159 -----GALSDRYIIAYNSLMDNFIMDVQSG-----YIDFGDGS 193
QY 209 NFGGLGYA-----TGAIHLDD-----TQGFRRKMGACQFSETNSICRHVLPKN 248
Db 194 PLNFFSYAGKNGHAYRSIGKVLIDRGEVKKEDMSQALRHGWHGTHSEAE--VRELLEQN 250

RESULT 7
PEDA_LACHE
ID PEDALACHE STANDARD; PRT; 474 AA.
AC Q48558; P71434;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dipeptidase A (EC 3.4.11.1).
GN PEDPA OR PEDP.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CMR232;
RX MEDLINE=96146518; PubMed=8550503;
RA Dudley E.G., Husgen A.C., He W., Steele J.L.;
RT "Sequencing, distribution, and inactivation of the dipeptidase A gene
(pepDA) from Lactobacillus helveticus CNR232.";
RL J. Bacteriol. 178:701-704(1996).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX STRAIN=53/7;
RX MEDLINE=96338998; PubMed=8766699;
RA Vesanto E., Peltonlehti K., Purtsi T., Steele J.L., Palva A.;
RT "Molecular characterization, over-expression and purification of a
novel dipeptidase from Lactobacillus helveticus";
RL Appl. Microbiol. Biotechnol. 45:638-645(1996).
CC -!- FUNCTION: HYDROLYSES A WIDE RANGE OF DIPEPTIDES BUT UNABLE TO
CC HYDROLYZE DIPEPTIDES CONTAINING PROLINE. HIGHEST ACTIVITY AGAINST
CC MET-ALA.
CC -!- CATALYTIC ACTIVITY: DIPEPTIDE + H(2)O = 2 AMINO ACID.
CC -!- ENZYME REGULATION: INHIBITED BY ZN(2+), CU(2+), CA(2+) AND CD(2+).
CC -!- SUBUNIT: HOMOCYMER (PROBABLE).
CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS OBSERVED AT PH 6.0 AND 55
CC DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY U34.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U34257; AAC43971.1; -.
CC EMBL; Z38063; CAA86210.1; -.
CC MEROPS; U34.001; -.
KW Hydrolase; Dipeptidase.
FT CONFLICT 211 211 T -> A (IN REF. 2).
SQ SEQUENCE 474 AA; 53512 MW; 3BFA79983D3CEEF2 CRC64;

Query Match 5.9%; Score 80.5; DB 1; Length 474;
Best Local Similarity 23.2%; Pred. No. 7.6;
Matches 39; Conservative 28; Mismatches 56; Indels 45; Gaps 9;
QY 43 DKIGDHHTHDEHSVSHVGLQAFETW-LQMHHATKQEV--VRYQAYLQSLGNLYLPMS 99
Db 301 DAYGDCQTPKQKTFPIGINRNFETHILQIRNDVPAETAGVQWLAFGPNFTNSMLPFYT 360

QY 100 QLTTARTSWOACGHEPYQLPPEHL---W-----GOIVPT-LHLYQDLKSRGILPANTOI 149
Db 361 NVTTTPEAWQT-----TPKFNLIKFWLNKLTAGLGDNTYRVYGELE-----402
QY 150 RSVYRNPELNOAGGAMSKHLTNSAIDIWVDPLEIKSQALYELQNLRL 197
Db 403 -DAFEOKSLAQC---HKTOHET-----DKEYVKNLSGRELQDKL 436

RESULT 8
CD5S_MOUSE
ID CD5S_MOUSE STANDARD; PRT; 369 AA.
AC Q35926; O35277;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cyclin-dependent kinase 5 activator 2 precursor (CDK5 activator 2)
DE (Cyclin-dependent kinase 5 regulatory subunit 2) (P39) (P391).
GN CDK5R2 OR NCKSAI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98322126; PubMed=9655938;
RA Nilden F., Baekstrom A., Bark C.;
RT "Molecular cloning and characterisation of a mouse gene encoding an
isoform of the neuronal cyclin-dependent kinase 5 (CDK5) activator.";
RL Biochim. Biophys. Acta 1398:371-376(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=129/SVJ;
RA Zheng M., Leung C.L., Liem R.K.H.;
RT "Comparative analysis of gene expression of the cyclin-dependent
Kinase 5 (cdk5) activators p35 and p39 in the rat CNS and embryonic
development revealed distinctive overlap with cdk5.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTIVATOR OF CDK5/TPKII.
CC -!- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A REGULATORY
CC SUBUNIT.
CC -!- SIMILARITY: BELONGS TO THE CYCLIN-DEPENDENT KINASE 5 ACTIVATOR
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U90267; AAC53595.1; -.
CC EMBL; AF016393; AAB69709.1; -.
DR MGD; MGI:1330828; Cdk5r2.
FT PROPEP 1 ?
FT CHAIN ? 369 ?
FT DOMAIN 75 78 POTENTIAL.
FT DOMAIN 155 163 CYCLIN-DEPENDENT KINASE 5 ACTIVATOR 2.
FT DOMAIN 348 351 POLY-LYS.
FT DOMAIN 351 351 POLY-PRO.
FT CONFLICT 91 91 G -> V (IN REF. 2).
FT CONFLICT 172 172 S -> G (IN REF. 2).
FT CONFLICT 191 191 R -> G (IN REF. 2).
FT CONFLICT 197 197 S -> R (IN REF. 2).
FT CONFLICT 220 220 L -> W (IN REF. 2).
FT CONFLICT 248 248 S -> R (IN REF. 2).
FT CONFLICT 253 253 A -> G (IN REF. 2).
SQ SEQUENCE 369 AA; 38914 MW; DA4EB1FD4D48CE66 CRC64;

Query Match 5.9%; Score 80; DB 1; Length 369;
Best Local Similarity 27.5%; Pred. No. 6.2;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:15:33 ; Search time 26.57 seconds
(without alignments)
1627.728 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 1355
Sequence: 1 MKNFNGYFTTLLISMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 375 | 27.7 | 240 | 2 Q93D15 | Q93D15 acinetobact |
| 2 | 122 | 9.0 | 562 | 16 Q92JX1 | Q92JX1 rhizobium m |
| 3 | 106.5 | 7.9 | 622 | 16 Q98H74 | Q98H74 rhizobium l |
| 4 | 102.5 | 7.6 | 523 | 16 Q989B8 | Q989B8 rhizobium l |
| 5 | 101 | 7.5 | 216 | 16 Q9A852 | Q9A852 caulobacter |
| 6 | 98.5 | 7.3 | 193 | 16 Q9A17 | Q9A17 rhizobium l |
| 7 | 90 | 6.6 | 539 | 5 Q95TM6 | Q95TM6 drosophila |
| 8 | 89.5 | 6.6 | 1377 | 5 Q9Y0X8 | Q9Y0X8 drosophila |
| 9 | 89.5 | 6.6 | 1377 | 5 Q9W2F3 | Q9W2F3 drosophila |
| 10 | 88 | 6.5 | 1004 | 16 Q84646 | Q84646 chlamydia t |
| 11 | 86 | 6.3 | 407 | 11 Q924J7 | Q924J7 spermophiliu |
| 12 | 86 | 6.3 | 536 | 16 Q07794 | Q07794 mycobacteri |
| 13 | 85.5 | 6.3 | 343 | 2 Q93CV9 | Q93CV9 lactobacill |
| 14 | 84.5 | 6.2 | 532 | 16 Q912X5 | Q912X5 pseudomonas |
| 15 | 84 | 6.2 | 543 | 2 Q9L868 | Q9L868 desulfovibr |
| 16 | 83.5 | 6.2 | 373 | 16 Q9KR92 | Q9KR92 vibrio chol |

ALIGNMENTS

RESULT 1

Q93D15 PRELIMINARY; PRT; 240 AA.
 AC Q93D15;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE HYPOTHETICAL 27.4 KDA PROTEIN.
 OS Acinetobacter sp. ADP1.
 OC Acinetobacter
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OX NCBI_TaxID=62977;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ADP1, BD413;
 RX MEDLINE=21555098; PubMed=11698371;
 RA Young D.M., Ornstun L.N.;
 RT "Functions of the Mismatch Repair Gene mutS from Acinetobacter sp.
 RT Strain ADP1.";
 RL J. Bacteriol. 183:6822-6831(2001).
 DR EMBL; AF400582; AAK92497.1; .
 KW Hypothetical protein.
 SQ SEQUENCE 240 AA; 27436 MW; E0FF4B7F5FFBC519 CRC64;

Query Match 27.7%; Score 375; DB 2; Length 240;
 Best Local Similarity 35.9%; Pred. No. 2.2e-28;
 Matches 89; Conservative 45; Mismatches 80; Indels 34; Gaps 10;

QY 11 TLISMLVACSAPIPTNQVSPKTPSVLITKDKIGD---HHTHEHDESYS-----H 59
 Db 12 SLIIPMMVGCT---TAPO-----KKGKILDLPGKHRIHIPOETVIERVPRK 56
 QY 60 VGLQAHFETWLOMHATKQEVVRYQAYLQSLRNLGVLPPMSOLLTTARSQACGHEPYOLP 119
 Db 57 VTPNSYL-NWLQ-NFSNRITQIEYQFLSRNDVANIVENYELLKTARDWQCGKSEYAVP 114
 QY 120 PHLNGQIVPTUHLVQDLKSRGILPANTQIRSVYRNPELNQACGAAMSKHLTNSAIDIW 179
 || :||| : : : || : : ||| : ||||| :||| |||||

Db 115 NRELWNAITPLRLVKYLVAAIL-TDFEVTSVYRDLPLNQAGGANSRHLFNSAIDFR 173
 QY 180 V-PDLEIKSQALYELQN---RLCYWLEHGENQFGLGYATGAIHLDTQGRKWAQFS 235
 Db 174 IGPEIP-QQDFAFIENTFKLCQFNQHQGLNGLGLYSSGQIHDTQGYRTWGPDS 232
 QY 236 EYNSICRH 243
 Db 233 RNSSMCNY 240
 RESULT 2
 Q92JX1 PRELIMINARY; PRT; 562 AA.
 AC Q92JX1;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN SMC04010.
 GN SMC04010.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21368234; PubMed=11474104;
 RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
 Barloy-Hubier F., Barnett M.J., Becker A., Boistard P., Botie G.,
 Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
 Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
 Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
 Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
 Kahn M.L., Kaiman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
 Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
 Ramsperger U., Surzycki K., Thebault P., Vandenbol M., Wong K., Yeh K.-C., Batut J.,
 Vorhoele F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
 RL Science 293:668-672(2001).
 DR EMBL; AL591791; CAC47373.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 562 AA; 59535 MW; 12C41231BACD3422 CRC64;

Query Match 9.0%; Score 122; DB 16; Length 562;
 Best Local Similarity 25.4%; Pred. No. 0.0024;
 Matches 43; Conservative 29; Mismatches 55; Indels 42; Gaps 10;

QY 74 HATKQEVVRYQAYLQSLRGNY-LPPMSQLTTARSQACGHEPYQLPPEHL---WQIYP 129
 Db 20 HTKEKAQITYK-----RNGRYDQKGLQINRFLRDWRR--NEPTKMDPRLLDLVW----- 67
 QY 130 TLHLQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLE 184
 Db 68 --EYQKSGSRDYI-----HVSAYRSPATNGMLRSKRGVAKSKHMLGKAMDFYIPDYK 121
 QY 185 IKSQALYELQNLQCYWLEHGENQFGLGYATGA---IHLDTQGRKW 230
 Db 122 LKT--LREVGKMF-----QVGGVGYPTSGSPFVHMDVGVRAW 158
 RESULT 3
 Q98H74 PRELIMINARY; PRT; 622 AA.
 AC Q98H74;
 DT 01-OCT-2001 (TREMELrel. 18, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMELrel. 18, Last annotation update)
 DE ML22999 PROTEIN.
 GN ML22999.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Phyllobacteriaceae; Mesorhizobium.
 OC NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003000; BAB49992.1; -
 KW Complete proteome.
 SQ SEQUENCE 622 AA; 65402 MW; 1E4C6B5F8C291655 CRC64;

Query Match 7.9%; Score 106.5; DB 16; Length 622;
 Best Local Similarity 23.4%; Pred. No. 0.086;
 Matches 40; Conservative 27; Mismatches 61; Indels 43; Gaps 9;

QY 72 MHATKQEVVRYQAYLQSLRGNYLPP-MSQLTTARSQACGHEPYQLPPEHL---WGQI 127
 Db 27 LHTHEKAEIV-----YKRNRYLPEGLRKINILRDWRR--NEPTKMDPRLLDLVW--- 75
 QY 128 VPTLHLQDLKSRGILPANTQIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDP 182
 Db 76 ---EAYRESGATDYI-----QVCGYRSPATNSMLRSKRGVAKSKHMLGKAMDFYIPG 127
 QY 183 LEIKSQALYELQNLQCYWLEHGENQFGLGYATGA---IHLDTQGRKW 230
 Db 128 VPLKRLRNIGL-----KMGGVGYPTSGSPFVHMDVGVNRHW 166
 RESULT 4
 Q989B8 PRELIMINARY; PRT; 523 AA.
 ID Q989B8
 AC Q989B8;
 DT 01-OCT-2001 (TREMELrel. 18, Created)
 DT 01-OCT-2001 (TREMELrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMELrel. 18, Last annotation update)
 DE MLR6494 PROTEIN.
 GN MLR6494.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OC NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099; PubMed=11214968;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003009; BAB52779.1; -
 KW Complete proteome.
 SQ SEQUENCE 523 AA; 55318 MW; FA6196E5665414C4 CRC64;

Query Match 7.6%; Score 102.5; DB 16; Length 523;
 Best Local Similarity 25.4%; Pred. No. 0.17;
 Matches 43; Conservative 29; Mismatches 58; Indels 39; Gaps 10;

QY 72 MHATKQEVV--RYQAYLQSLRGNYLPPMSQLTTARSQACGHEPYQLPPEHLWGQIYP 129
 Db 28 LHTGEKAEIVKRNRYDQAGL-----KKIDFMLRDWRR--NEPTKMDPRLL--DLV- 75

QY 130 TLHLQDLKSRGILPANTOIRSVYRNPELN-----QCAGGAAMSKHLTNSAIDIWVDPLE 184
 Db 76 ----WQAYRASG--SSAYTHVVSAYSPATNMLRSKGVARESOHWVGRAMDFFLPDVP 130
 QY 185 IKSQALYELQNLKCOYLWLEHGENQFGLGLYATGA---IHLDTQGFGRKW 230
 Db 131 LKK--LRDIGLKM-----QGGGVGYPTSGSPFIHMDVGNVRHW 167

RESULT 5
 Q9A852 PRELIMINARY; PRT; 216 AA.
 ID Q9A852
 AC Q9A852;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN CC1512.
 GN CC1512.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 CC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Eisen J., Feldblyum T.V., Laub M.R.T., Paulsen I.T., Nelson K.E.,
 RA Eilen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A.G., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kollonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Iran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005826; AAK23491.1;
 DR TIGR; CC1512;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 216 AA; 23597 MW; 65D0901102F7BBD4 CRC64;

Query Match 7.5%; Score 101; DB 16; Length 216;
 Best Local Similarity 22.7%; Pred. No. 0.073;
 Matches 50; Conservative 29; Mismatches 79; Indels 62; Gaps 10;

QY 28 PQVSPKTPSVLIT-----KKIGDHTTHEHDSVSHVGLQAHFETWLMHATK 77
 Db 39 PQATPVPPPTVAATVASIDPPALKPAVDPWWHLN-----VHTGEK 81

QY 78 QEVRYQAYLQSLGNLTP-PMQLLTARSQACGHEPVQLPPEHLWGQIVPTLHLYQD 136
 Db 82 LEAVYWN-----GDYVPAVSALDKVLRDYNDEVHPIDRGLYDLLDQIA-----RK 129

QY 137 LKSRGILPANTOIRSVYRNPELNQAGG-----AAMSKHLTNSAIDIWVDPLEIKSQALY 191
 Db 130 TQSKGPF-----QVIGSVSPATNLLSKRSGEVAKSLHMDGKAMDIFLEDVCLKHVRAA 185

QY 192 ELQNLKCOYLWLEHGENQFGLYAT-GAHLDTQGFGRKW 230
 Db 186 ALD-----LSVG-----QGVGYPTSNFVHVDVGPVRKW 213

RESULT 6
 Q98A17 PRELIMINARY; PRT; 193 AA.
 ID Q98A17
 AC Q98A17;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 DE MLL5985 PROTEIN.
 GN MLL5985.
 OS Rhizobium loti (Mesorhizobium loti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003008; BAB52343.1;
 DR InterPro; IPR00182; Acetyltransf_GCN5.
 DR Pfam; PF00583; Acetyltransf; 1.
 KW Complete proteome.
 SQ SEQUENCE 193 AA; 21306 MW; DA78C07070694B5E CRC64;

Query Match 7.3%; Score 98.5; DB 16; Length 193;
 Best Local Similarity 21.2%; Pred. No. 0.11; Mismatches 27; Indels 41; Gaps 6;
 Matches 38; Conservative 27;

QY 54 DESVSHVGL--QAHFETW-----LQMHATKQEVRYQAYLQSLGNLYPPMSQLLTT 104
 Db 25 DATVADVDLIAQLHIESWRNAYAGILSAGPIVLDRTQTVWRERFNSPAPNL-KVIVA 83

QY 105 ARSQACGHEPVQLPPEHLWGQIVPTLHLYQDLKSRGILPANTOIRSVYRNPELNQAGG 164
 Db 84 EREGQALGFVCLLGNADRWGTLVNLHVLPTAKRGV-----GRHLIRVAAG 131

QY 165 AAMSKHLTNSAIDIWVDPLEIKSQALY-----LQNLKCOYLWLEHG 205
 Db 132 WS-AENYPGVGLHWLVYEVNAPARAFYERMGGQVVARLPQSNPDGRIHVELCYYPDGS 189

RESULT 7
 Q95TM6 PRELIMINARY; PRT; 539 AA.
 ID Q95TM6
 AC Q95TM6;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE LD38816P.
 GN CG8793.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y, CN BW SP;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY058679; AAL13908.1;
 SQ SEQUENCE 539 AA; 59662 MW; 423555B44D084EFA CRC64;

Query Match 6.6%; Score 90; DB 5; Length 539;
 Best Local Similarity 27.2%; Pred. No. 2.8;
 Matches 41; Conservative 19; Mismatches 63; Indels 28; Gaps 6;

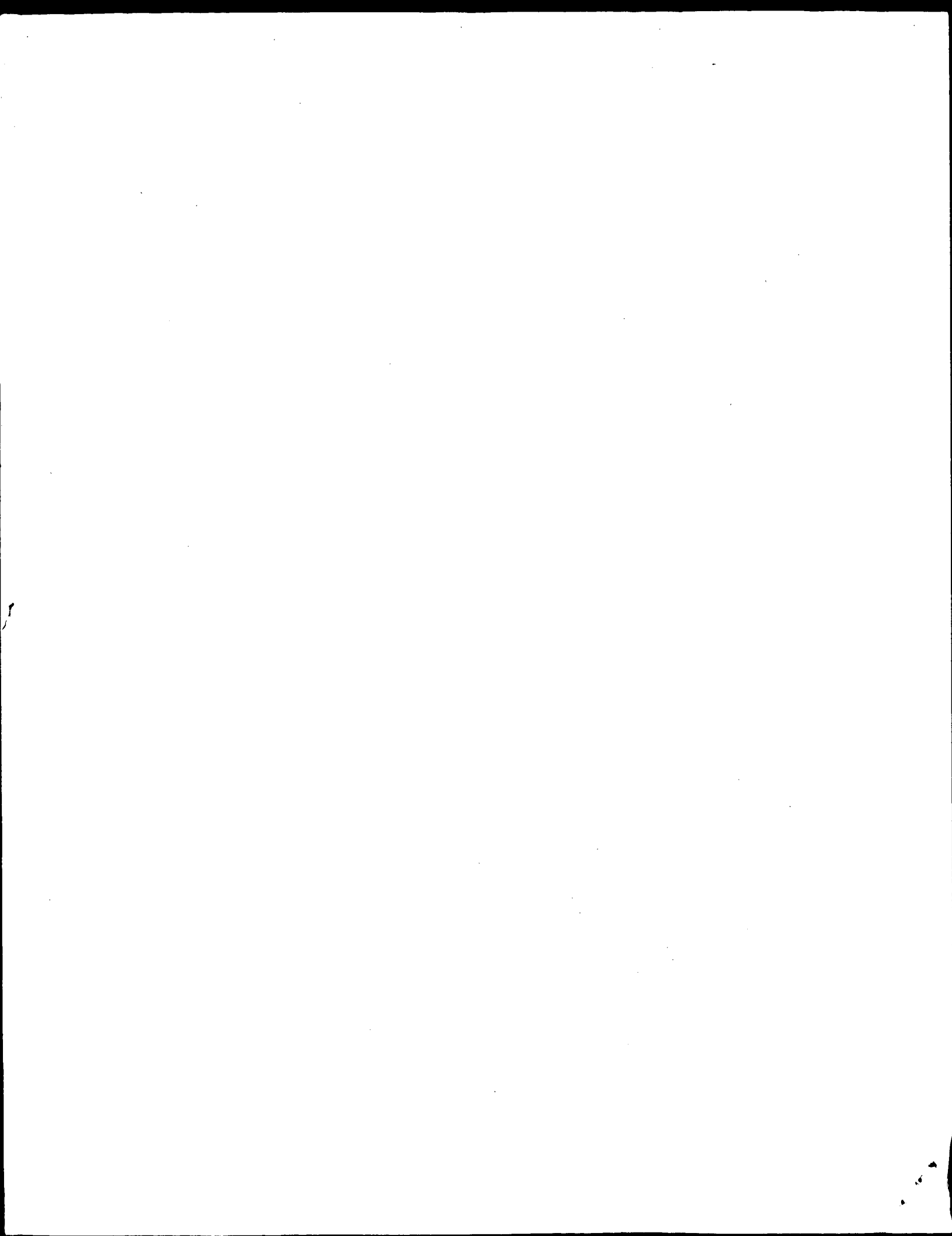
QY 54 DESVSHVGLQAHFETWLMHATKQEVRYQAYLQSLGNLYPPMSQLLTTARSQACGH 113
 Db 240 DPAASLLG-TLOFETQKTRPHNAKQS---SQTVMNRLTIKVLQPLPAMNYSFT----- 289

| | | | | |
|-----------------------|------------------|----------------|------------|--------------|
| Query Match | 6.6%; | Score 89.5; | DB 5; | Length 1377; |
| Best Local Similarity | 18.8%; | pred. No. 11; | | |
| Mismatches | 51. Conservative | 42. Mismatches | 87: Indels | 91: Gaps |
| | | | | 10; |

[illegible]

QY 75 ATKQEVRYQAYLQSRNGNYLPPM-----SQLTTARSWOACGHEPYQLPPEHLWGQI 127

Search completed: July 30, 2002, 16:19:11
Job time: 218 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:15:58 ; Search time 29.76 Seconds
(without alignments)
933.080 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFITTLSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-----------------------------|
| 1 | 250 | 100.0 | 250 | 22 | AAB60645 Moraxella catarrha |

ALIGNMENTS

RESULT 1
AAB60645
ID AAB60645 standard; Protein; 250 AA.
XX
AC AAB60645;

XX DT 04-MAY-2001 (first entry)
XX DE Moraxella catarrhalis strain ATCC43617 BASB120 protein.
XX KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
XX KW genetic immunisation; infection; upper respiratory tract; otitis media;
XX KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
XX KW invasive disease; antibacterial; auditory.
XX OS Moraxella catarrhalis.
XX PN WO200109335-A2.
XX PD 08-FEB-2001.
XX PF 31-JUL-2000; 2000WO-EP07361.
XX PR 03-AUG-1999; 99GB-0018281.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Thonnard J;
XX DR WPI: 2001-159872/16.
XX DR N-PSDB; AAF59797.
XX PT New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis strain American Type Culture Collection 43617, for use as therapeutic agents or vaccines against bacterial infections, e.g. otitis media or pneumonia
XX PS Claim 4; Page 64; 75pp; English.
XX CC The invention relates to the Moraxella catarrhalis strain ATCC43617
XX CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
XX CC invention also relates to immunogenic fragments of the BASB120 protein,
XX CC expression vectors and host cells comprising BASB120 nucleic acids, the
XX CC recombinant production of BASB120, vaccine compositions comprising the
XX CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
XX CC compositions comprising the anti-BASB120 antibody, and a method of
XX CC identifying a Moraxella catarrhalis infection via the detection of
XX CC BASB120 proteins or antibodies. The vaccine compositions of the invention
XX CC are useful as prophylactic or therapeutic agents against Moraxella
XX CC catarrhalis infections in mammals, particularly humans. Moraxella
XX CC catarrhalis is a Gram negative bacterium frequently isolated from the
XX CC human upper respiratory tract, which is responsible for several
XX CC pathological conditions. It is responsible for about 15% of otitis media
XX CC cases in children (which can lead to temporary or permanent hearing
XX CC loss). It also causes pneumonia in elderly people, and sinusitis,
XX CC nosocomial infections and, less frequently, invasive diseases. BASB120
XX CC proteins or nucleotides may additionally be used in screening for novel
XX CC antibacterial compounds, and in the diagnosis and staging of infections.
XX CC The present sequence represents the Moraxella catarrhalis strain
XX CC ATCC43617 BASB120 protein.
XX SQ Sequence 250 AA;
Query Match 100.0%; Score 250; DB 22; Length 250;
Best Local Similarity 100.0%; Pred. No. 1.3e-258;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKNFNQYFITTLSMLVACAPIPTNPQVSPKTPSVLLTKDKIGDHHHDESHV 60
DB 1 mknfnqyfitlissmlvacapiptnpqvspiktpsvlltkdkigdhthdeesvshv 60
QY 61 GLQAHFETWLOMHATKQEVVRYQAYLQSLGNLPPMSQLLTARTSWQACGHEPYQLPP 120
DB 61 glqahfetwloqmhbatkqevvryqaylqsrlnylnppmsqlttrswqacghepyqlpp 120
QY 121 EHLNGQIVPTLHLLODLKSRGILPANTQIRSVYRNPELNQCAGAAKSHLTNSAIDIWY 180
DB 121 ehlngqivptlhlldklsrgilpantqirsvyrnpelnqcagaaakshltnsaidiwy 180

Db 121 ehlwggivptlhlyqdlksrgilpantqirsvyrnpelnqcaggaamskhltnsaidiww 180
Qy 181 PDLETKSQALYELQNRLCQYWLEHGENONFGLGLYATGAIHLDTQGPWKGAOFSETNSI 240
Db 181 pdleiksqalyelqnrlicqywlehgengnfglglyatgaihlldtqgfrkvgagfsetnsi 240
Qy 241 CRHVLPKNKL 250
Db 241 crhvlpknkl 250

Search completed: July 30, 2002, 16:19:45
Job time: 227 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:17:53 ; Search time 13.06 Seconds
(without alignments)
467.565 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFITTLLISSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

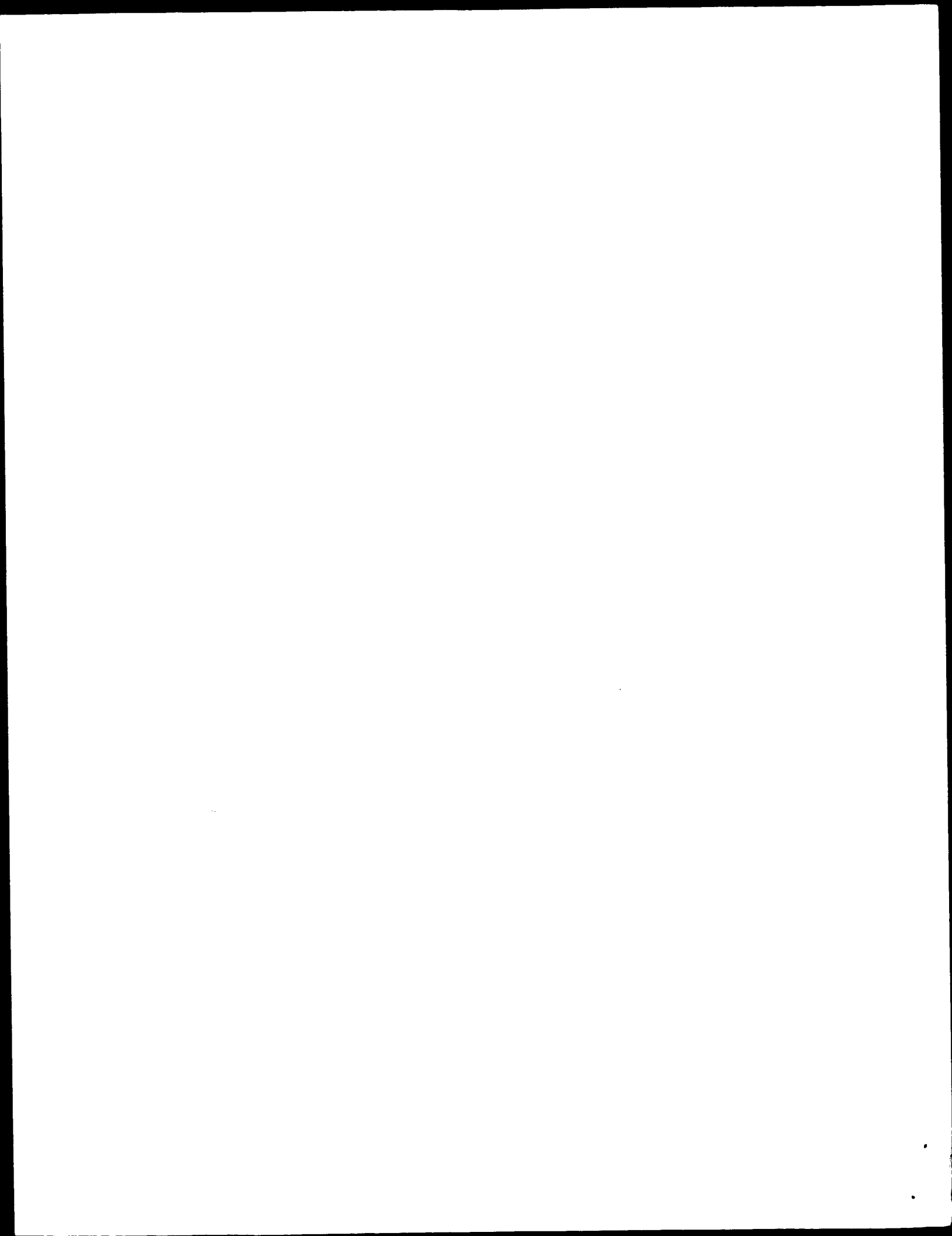
Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| Result | Query | % | Score | Match | Length | DB | ID | Description |
|------------------|-------|---|-------|-------|--------|----|----|-------------|
| SUMMARIES | | | | | | | | |
| ----- | | | | | | | | |
| No matches found | | | | | | | | |

Search completed: July 30, 2002, 16:20:05
Job time: 132 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:18:23 ; Search time 19.33 Seconds
(without alignments)
1242.749 Million cell updates/sec

Title: US-09-674-779-2

Perfect score: 250

Sequence: 1 MNFNQYFITTLLISSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table:

OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

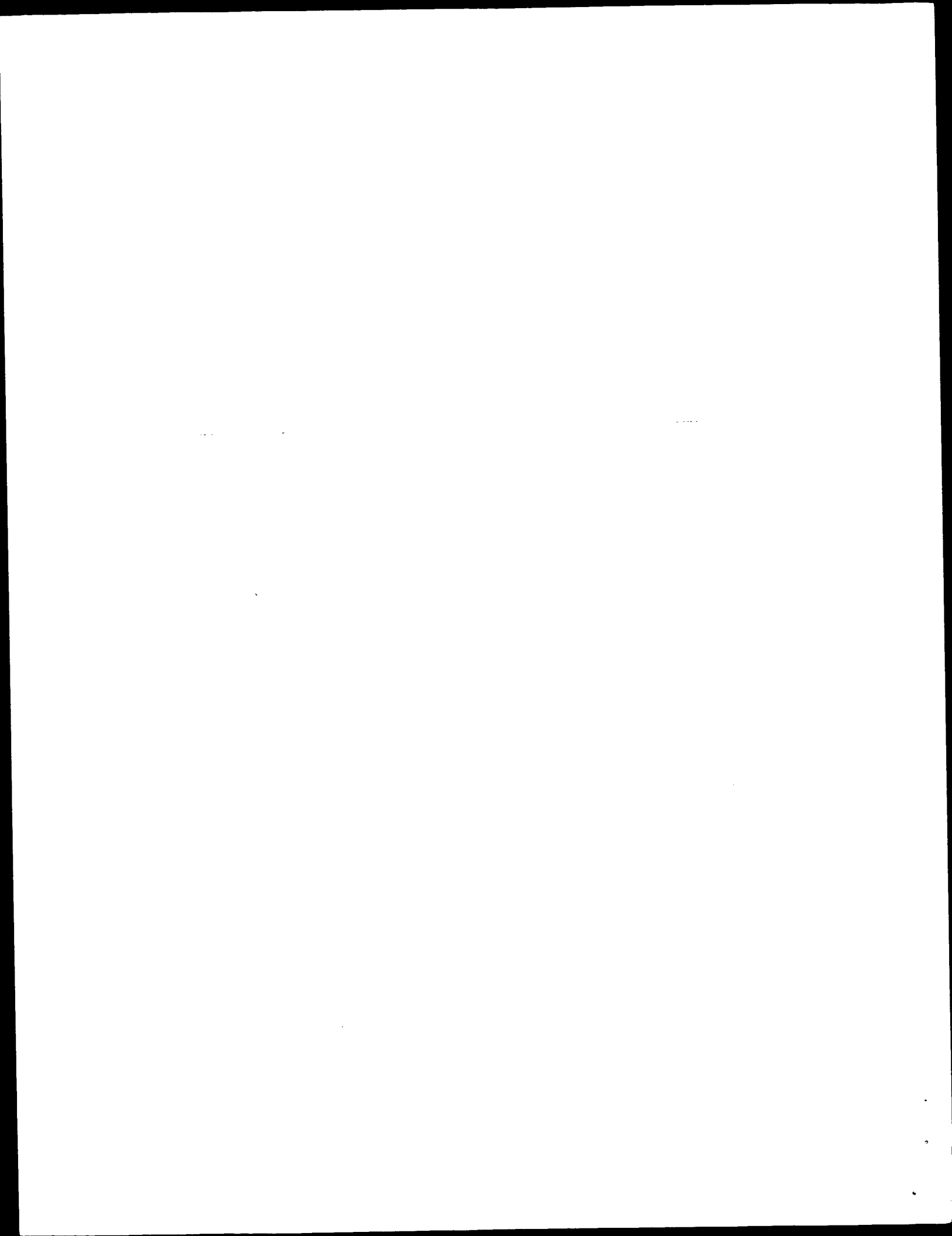
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Score | Match | Length | DB | ID | Description |
|---------------|----------------|-------|--------|----|----|-------------|
|---------------|----------------|-------|--------|----|----|-------------|

No matches found

Search completed: July 30, 2002, 16:20:31
Job time: 128 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:20:08 ; Search time 10.97 Seconds
(without alignments)
882.396 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFNQYFITTLSMLVAC.....GAQFSETNSICRHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 105224 seqs, 38719550 residues
Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

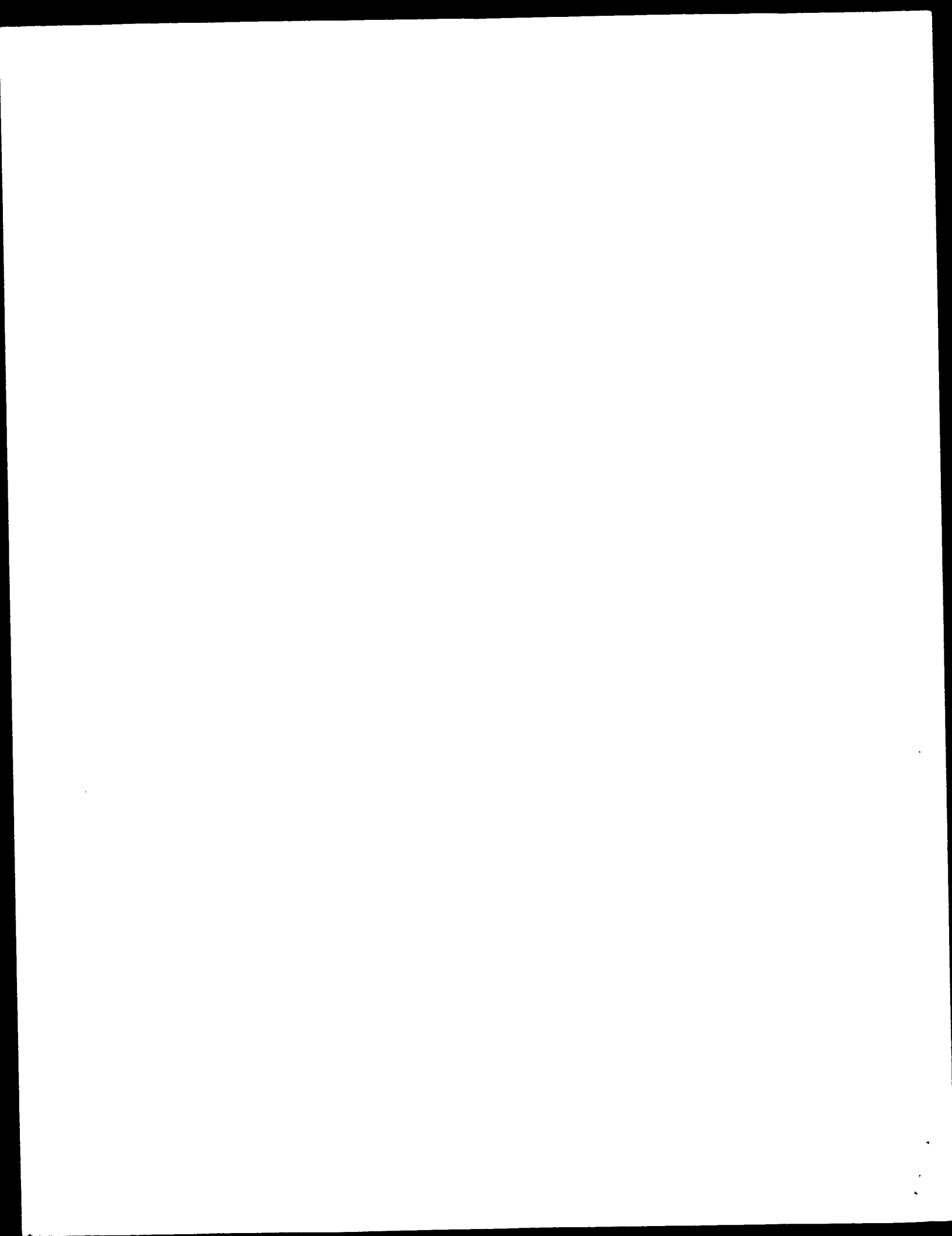
Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| SUMMARIES | | | |
|------------------|-------|--------------|-------------------|
| Result | Query | | |
| No. | Score | Match Length | DB ID Description |
| ----- | ----- | ----- | ----- |
| No matches found | | | |

Search completed: July 30, 2002, 16:23:47
Job time: 219 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 16:19:48 ; Search time 26.55 seconds
(without alignments)
1628.954 Million cell updates/sec

Title: US-09-674-779-2
Perfect score: 250
Sequence: 1 MKNFQYFITTLLSSMLVAC.....GAQFSETNSICHHVLPKNKL 250

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|---------------|-------|----------------|--------|----|-------------|
|---------------|-------|----------------|--------|----|-------------|

No matches found

Search completed: July 30, 2002, 16:23:30
Job time: 222 sec

